Technical Report No. 57

Analytical Method Validation and Transfer for Biotechnology Products

2012

PDA Analytical Method Validation And Transfer For Biotechnology Products Task Force Members

Authors

Stephan O. Krause, Ph.D., Chair, MedImmune

Florence Baudoux, Ph.D., GlaxoSmithKline

Pierre Douette, Ph.D., Eurogentec S.A.

Nicole Gibelin, Ph.D., Sanofi Pasteur, France

Alice E. Grebanier, Ph.D., Centocor R&D

Rajesh Krishnamurthy, Ph.D., Zyngenia, Inc.

Carl-Gustav, Millinger, Ph.D., Swedish Orphan Biovitrum

Frank Moffatt, Ph.D., Solvias AG

Dwayne Neal, SAIC-Frederick, Inc.

Phillip Ramsey, SAIC-Frederick, Inc.

Michael Rooney, Ph.D., BioTechLogic

Melissa J. Smith, MJ Quality Solutions

Michael Warncke, Ph.D., Bayer HealthCare

Earl K. Zablackis, Ph.D., Sanofi Pasteur, USA

Contributors

Patricia W. Cash, Ph.D., MedImmune

Larissa Chirkova, Novozymes

Marta Germano, Pharming Technologies

Siegfried Giess, Ph.D., Paul-Ehrlich Institute

Rashmi Rawat, Ph.D., U.S. FDA

Analytical Method Validation and Transfer for Biotechnology Products

Technical Report No. 57

ISBN: 978-0-939459-41-4

Table of Contents

FIGURES AND TABLES INDEX

1.0 Introduction

This Technical Report (TR) provides risk-based guidance for Analytical Method Validation (AMV), which follows Analytical Method Development (AMD) or Analytical Method Qualification (AMQ), and contains risk-based guidance for other, related method lifecyle steps, such as Analytical Method Transfer (AMT).

The guidance provided here builds upon the International Conference on Harmonization (ICH) Q2 (R1) guidelines and includes additional considerations for analytical platform technology (APT) methods as well as the impact of stakeholder considerations, and essentially all modern quality expectations as recommended in the ICH Q8 (R2), Q9, and Q10 guidelines ***(1–4).***

Similar to the manufacturing process, an analytical method can also be considered to be a process. The validation strategy for analytical methods could therefore conceptually follow those of Process Validation ***(5).*** AMV can then be defined as the collection and evaluation of data, from the analytical method development stage throughout routine QC testing, which establishes scientific evidence that an analytical method is capable of consistently delivering accurate and reliable results.

1.1 Scope and Purpose

This TR is to provide practical and strategic guidance to efficiently use historical data and knowledge to design suitable risk-based AMV studies, and set appropriate protocol acceptance criteria. The typical method lifecycle steps prior, during, and beyond the AMV studies are illustrated in **Figure 1.1-1**. The typical steps prior to validation, usually performed at early pharmaceutical development stages, are included in this figure to show the dependency among early- and late-stage lifecycle steps. The AMV process begins with the validation readiness assessment and continues with the post-validation steps, maintenance (validation continuum), transfer(s), comparability, as they may apply to the continuous demonstration of analytical method suitability. The typical sequence of all prevalidation, validation and post-validation steps, as illustrated in the bottom half of **Figure 1.1-1,** is reflected in the sequence of sections in this TR. Instead of dealing in great detail with many possible exceptions and special considerations, this TR is intended to provide practical guidance to typical development processes and AMV studies.

The guidance presented in this TR applies to all biotechnological manufacturers and all contract development and manufacturing organizations. This TR does not provide specific guidance for the timing of AMV study execution with respect to the parallel product development lifecycle stages or guidance for analytical instrument qualification.

It should be considered that various new analytical technologies and/or the use of Process Analytical Technology (PAT) methods may suggest some modification to the validation strategies presented here. Specific aspects for the validation of bioassays such as curve fitting models and statistical reference-to-sample parallelism requirements are not covered in this TR. Case-specific considerations for microbiological method validation such as statistical sampling and testing environment conditions are also not covered as they depend on the analytical methodology and the intended use.

AMV studies are typically executed for future routine-use methods but may not be required for analytical methods used in support of pharmaceutical development ***(5).*** **Figure 1.1-2** illustrates the two different analytical method lifecycle paths separated according to the intended use of a particular method. The intended use of a particular method can be assessed early as part of the overall quality target product profile (QTPP) and a method should be selected accordingly. The intended use should be further considered when developing, qualifying and validating analytical methods. For example, measuring a critical quality attribute (CQA) or a critical process parameter (CPP) may require a more rigorous approach to the overall validation process. The intended use of a method can change during the method and/or product lifecycle(s) due to a specification change or other reasons.

Figure 1.1-1 Analytical Method Lifecycle Steps from Selection to Qualification or Validation

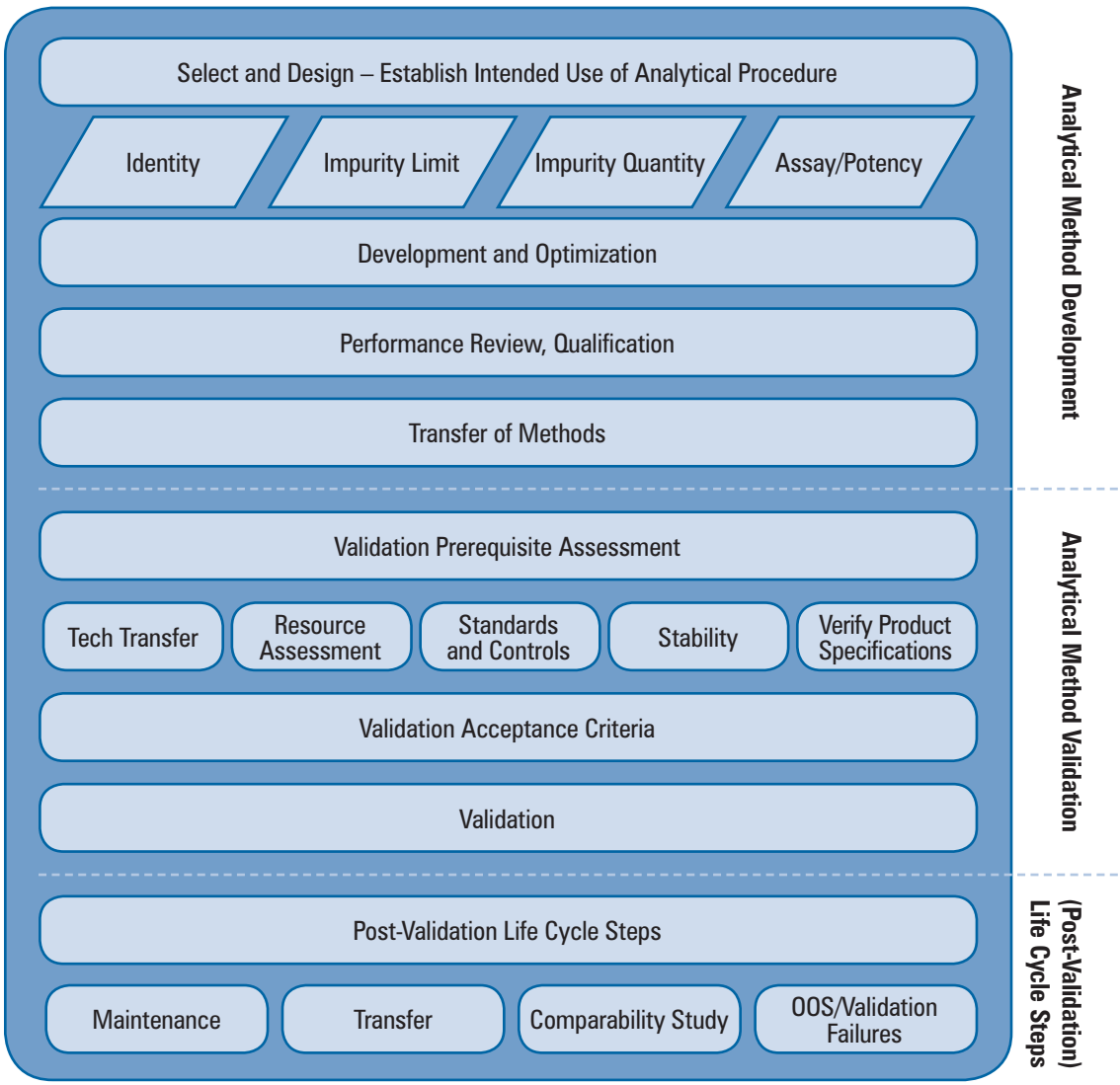
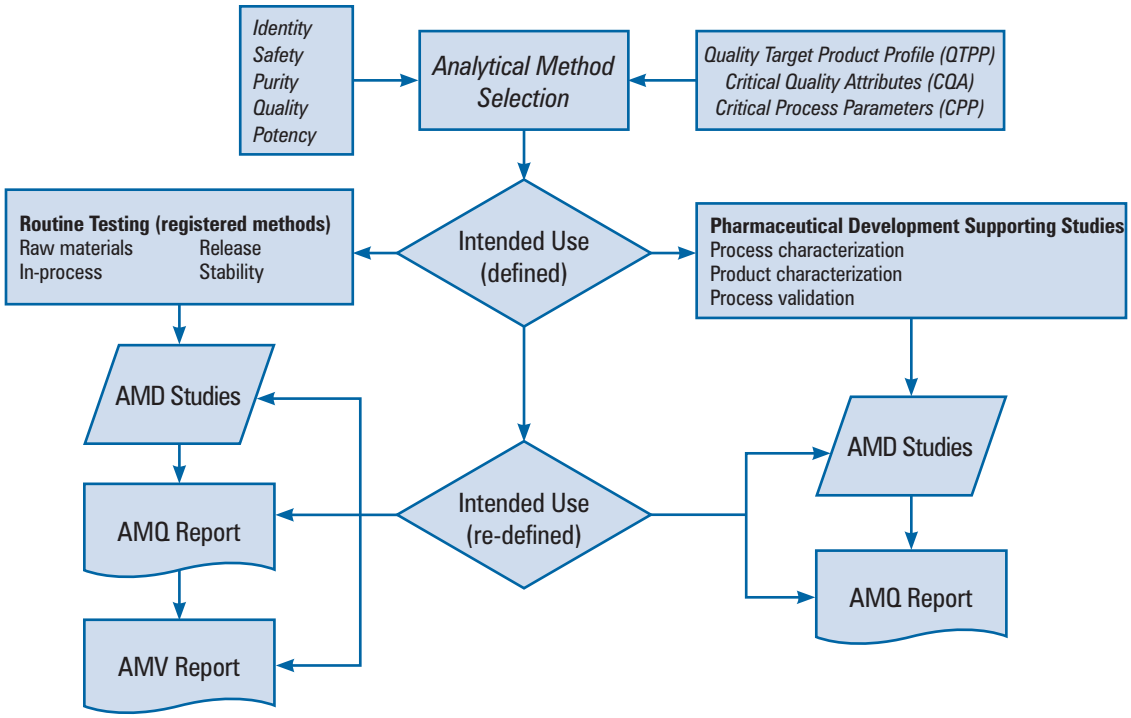


Figure 1.1-2 Example of a Method Lifecycle from the Identification of the Intended Use to Post-Validation Maintenance



2.0 Glossary of Terms

Acceptance Criteria
Numerical limits, ranges, or other suitable measures for acceptance of the results of analytical method validation that is satisfied to determine suitability of test method performance *(6)*.

Accuracy
An analytical procedure expresses the closeness of agreement between the value that is accepted either as a conventional true value or an accepted reference value and the value found. This is sometimes termed trueness *(1)*.

Analysis of Variance (ANOVA)
A general statistical approach to data analysis (i.e., comparison of means) in which the variation in a method's results is partitioned among explanatory factors in order to systematically assess factor influence and/or variance components.

Analyte
A specific chemical moiety being measured, which can be intact drug, biomolecule or its derivative, impurity, and/or excipients in a drug product *(7)*. [Synonym: measurand]

Analytical Instrument Qualification (AIQ)
The qualification of the analytical instrument(s) used as part of the analytical procedure.

Analytical Platform Technology (APT)
An analytical method that is used for multiple products and/or types of sample matrix without modification of the procedure. Similar to compendial methods, an APT method may not require full validation for each new product or sample type.

Analytical Procedure
That which is performed in order to obtain a reportable result. The procedure should describe in detail the steps necessary to perform the analytical test. This may include but is not limited to: the sample, the reference standard and the reagents preparations, use of the apparatus, generations of the calibration curve, use of the formulae for the calculation *(1)*. [Synonym: Method, Assay]

Bias
A systematic difference in a method that manifests itself as a deviation of the method mean from an expected value.

Biological Activity
Property that describes the specific ability or capacity of a product to achieve a defined biological effect *(8)*.

Bioanalytical Test Method
A method used to assess the presence of analytes (chemical or biological) in biological samples (e.g., serum, plasma, etc.) *(7)*.

Bioassay
Analysis (as of a drug) to quantify the biological activity(ies) of one or more components by determining its capacity for producing an expected biological activity.

Blocking
The grouping of related experimental units used in design of experiments (DOE).

Calibration Curve
The relationship between measured response values and analytical concentrations of a standard or reference material.

Coefficient of Determination (r^2)
A measure of the proportion of the variation of one variable determined by the variation of the other.

Comparability, Method
The demonstration of analytical method comparability (AMC) for method replacements.

A study to demonstrate that a modification to an existing method either improves or does not significantly change the analytical procedure's characteristics relative to the methods' validation and intended use.

Compendial Procedure
A method that is considered validated as published in one of the recognized compendia.

Confidence Interval
An interval estimate (range of values) of a population parameter, calculated from a random sample of the underlying population.

Correlation Coefficient (r)
A measure of covariation, the square root of the coefficient of determination.

Co-Validation
Sending and receiving laboratories participate in the AMV study execution.

Critical Reagent
A component of the test method that may have a substantial impact on the consistency and reliability of method performance. Features of critical reagents include:

1. A reagent that requires qualification of each new batch prior to routine use in an analytical procedure, or
2. A material whose method performance characteristics may change over time, during handling, or from lot to lot.
3. An analytical reagent that may be purchased only from a single vendor.

Reagent Examples: antibodies or enzymes that require titration prior to use, tissue culture treated plates when only one vendor's plates give acceptable results for a bioassay, growth factors for bioassay cells, conjugated proteins that require custom preparations, or reference or system suitability standards.

Degradation Product
Molecular variants resulting from changes in the desired product or product-related substance brought about over time and/or by the action of light, temperature, pH, water, etc., or by reaction with an excipient and/or the immediate container/closure system. Such changes may occur because of manufacture and/or storage (e.g., deamidation, oxidation, aggregation, proteolysis). Degradation products may be either product-related substance or product-related impurities ***(8).***

Design of Experiments (DOE)
A structured, organized method for determining the relationship between factors affecting an assay and output of that assay ***(2).***

Design, Experimental
The arrangement of factors and factor levels. Optimal experimental design minimizes "noise" in data to allow focus on the influence on assay response of critical factors. A factorial experiment (DOE) may minimize experiments required to achieve analytical purpose. (May be modified with complete block, factorial, fractional factorial, full factorial, incomplete block) ***(9).***

Drug Product
A pharmaceutical product type that contains a drug substance, generally, in association with excipients ***(8).*** [**Synonym:** Dosage Form; Finished Product]

Drug Substance
The active ingredient that is subsequently formulated with excipients to produce the drug product. It can be composed of the desired product, product-related substances, and product- and process-related impurities. It may also contain excipients, including other components such as buffers ***(8).*** (Synonyms: bulk drug substance, bulk material)

Equivalence
A comparison with the primary objective of showing that the results from two methods differ by an amount which has negligible impact on fitness for use. This is usually demonstrated by showing that the true difference is likely to lie between a lower and an upper equivalence margin of acceptance differences ***(10).***

Equivalence Margin
The largest difference between the results from two methods that is considered as being scientifically and statistically acceptable. ***(10).***

Error
Deviation from expected value. Error may be random or systematic.

Excipient
An ingredient added intentionally to the drug substance that should not have pharmacological properties in the quantity used ***(8).***

Factor
Independent variables that may influence assay outcome. (May be modified with confounded, crossed, fixed, interaction, level, modifying, nested, random) ***(9).***

Identification
Use of an analytical procedure to determine the chemical and biochemical identity of a material.

Impurity
Any component present in the drug substance or drug product that is not the desired product, a product-related substance, or excipient including buffer components. It may be either process- or product-related ***(8).***

Independent Replicates
Two or more measurements or observations that are generated from independently prepared samples and do not affect each other.

Limit, Detection (DL)
The lowest amount of analyte in a sample that can be detected but not necessarily quantitated as an exact value by an individual analytical procedure ***(1).*** [Synonym: Limit of detection (LOD)]

Limit, Quantitation (QL)
The lowest amount of analyte is a sample that can be quantitatively determined with suitable precision and accuracy by an individual analytical procedure ***(1).*** [Synonym: Limit of quantitation (LOQ)]

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

Matrix
The combination of materials (e.g., excipients, stabilizer components, etc.) which are components together with the measured analyte.

Matrix Effect
The direct or indirect alteration or interference in response due to the presence of additional sample components due to sample preparation (for analysis) or other interfering substances in the sample (product related excipients or residuals) ***(7).***

Method Capability
The resulting acceptable uncertainty of results to achieve the required capability to detect, quantify, and/or discriminate the analyte at levels that is relevant to the intended use.

Method Development
A process that involves the selection, optimization, and qualification of a physical/chemical, biological, molecular, or microbiological test procedure.

Method Qualification
Experimental studies performed to confirm the inherent performance capabilities of a test method for the material being analyzed and the intended use of the method. Method qualification can be performed during early development phases, prior to method validation. Specific method qualification characteristics (e.g., accuracy, specificity) should be confirmed based on the intended use of the analytical method and the relevant risk(s).

Method Validation
A formal, archived demonstration of the analytical capacity of an assay that provides justification for use of the assay for an intended purpose. Validations are conducted prospectively according to a written, approved plan that states acceptance criteria.

Method, Qualitative
An analytical procedure, based on the characteristics of a material that yields results that are not amenable to reliable enumeration.

Method, Quantitative
An analytical procedure that yields numerical results compared to quantitative specification(s).

Noninferiority
A comparison with the primary objective of showing that the result from one method is not inferior to the method being compared. This is usually demonstrated by showing that the true difference is likely to lie above the lower equivalence margin ***(10).***

Potency
The measure of the biological activity using a suitably quantitative biological assay, based on the attribute of the product that is linked to the relevant biological properties.

Precision
The closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions. Precision may be considered at three levels: repeatability, intermediate precision, and reproducibility. It is usually expressed as the variance, standard deviation, or coefficient of variation of a series of measurements *(1)*.

Precision, Repeatability
The closeness of agreement between a series of measurements obtained under ideal conditions (e.g., same day, analyst, and instrument) *(1)*.

Precision, Intermediate
The closeness of agreement between a series of measurements obtained within laboratory variations (e.g., different days, different analysts, different equipment) *(1)*.

Precision, Reproducibility
The closeness of agreement between a series of measurements for the same sample obtained among different laboratories *(1)*.

Quality
The degree to which a set of inherent properties of a product, system or process fulfills requirements *(3)*.

Quality Attribute
A molecular or product characteristic that is selected for its ability to help indicate the quality of the product, such as identity, purity, potency stability and safety *(11)*.

Quality Attribute, Critical (CQA)
Physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality. CQAs are generally associated with the drug substance, excipients, intermediates, and drug product *(2)*.

Quality Risk Management
A systematic process for the assessment, control, communication, and review of risks to the quality of the drug (medicinal) product across the product lifecycle *(3)*.

Range
The range of an analytical procedure is the interval between the lower and upper quantitation limits. Within this range, a suitable performance level for precision, accuracy, and linearity can be demonstrated *(1)*.

Reagent
For analytical procedures, any substance used in a reaction for the purpose of detecting, measuring, examining, or analyzing other substances.

Recovery
A measure of the amount of analyte carried through the entire sample preparation and assay procedure and expressed as a percentage of the nominal concentration.

Reference Standard
The defining characteristics of a reference standard are: 1) it is stable; 2) it performs similarly (e.g., on dilution) to test materials in the assay; and 3) it is homogeneous.

Regression
A mathematical model in which the response of a dependent variable is a function of change in an independent variable, such as is seen in a concentration-response curve. Regression may be linear (e.g., a straight line) or non-linear (e.g., four parameter logistic).

Repeatability
The precision under the same operating conditions over a short interval of time.

Replicates
Independent preparations of a sample or standard that are subject to the same treatment conditions *(12)*.

Reportable result
The final analytical result. This result is defined in the written approved test method and derived from one full execution of that method, starting from the original sample *(12)*.

Reproducibility
The precision among multiple laboratories (collaborative studies, usually applied to standardization of methodology) *(1)*.

Robustness
The measure of capacity to remain unaffected by small, but deliberate, variations in method parameters and provides an indication of its reliability during normal usage ***(1).***

Specification
A list of tests, references to analytical procedures, and appropriate acceptance criteria that are numerical limits, ranges, or other criteria for the test described. It establishes the set of criteria to which a material should conform to be considered acceptable for its intended use. Drug product and drug substance specifications are critical quality standards that are proposed and justified by the manufacturer and approved by regulatory authorities ***(6, 8).***

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

Spiking
The addition of a small known amount of a known compound to a standard, sample, or placebo, typically for the purpose of confirming the performance of an analytical procedure or the calibration of an instrument ***(7).***

Stability
The chemical/biological fidelity of an analyte in a given solvent/matrix under specific conditions.

Stability-indicating analytical methods
A test procedure that is able to discern changes in an analyte due to degradation processes. It is capable of accurately measuring changes in the product that can occur under conditions of physical or chemical stress.

Standard Deviation
The statistical measure of the dispersion of the data.

State of Control
A condition in which the set of controls consistently provides assurance of continued process performance and product quality ***(4).***

Suitability, system
Acceptance criteria for a valid reported result(s). [Synonym: Assay quality control]

Superiority
A comparison with the primary objective of showing that the result from one method is superior to the method being compared. This is usually demonstrated by showing that the true difference is likely to lie between zero and the upper equivalence margin ***(10).***

Validation
A documented program that provides a high degree of assurance that a specific process, method, or system will consistently produce a result meeting predetermined acceptance criteria ***(13).***

Validation, partial
A documented prospective study intended to demonstrate suitability for the intended use of previously validated methods, specifically for new products and/or processes ***(7).***

Validation, process
The collection and evaluation of data, from the process design stage through commercial production, which establishes scientific evidence that a process is capable of consistently delivering quality products ***(5).***

2.1 List of Abbreviations

AMC: Analytical Method Comparability

AMD: Analytical Method Development

AMM: Analytical Method Maintenance

AMQ: Analytical Method Qualification

AMT: Analytical Method Transfer

AMV: Analytical Method Validation

ANOVA: Analysis of Variance

APT: Analytical Platform Technology

CV: Coefficient of Variation

CI: Confidence Interval

CPP: Critical Process Parameter

CQA: Critical Quality Attribute

DL, LOD: Detection Limit, Limit of Detection

DOE: Design of Experiments

QL, LOQ: Quantitation Limit, Limit of Quantitation

QTPP: Quality Target Product Profile

SD: Standard Deviation

SOP: Standard Operating Procedure

TOST: Two One-Sided T tests

TR: Technical Report

VMP: Validation Master Plan

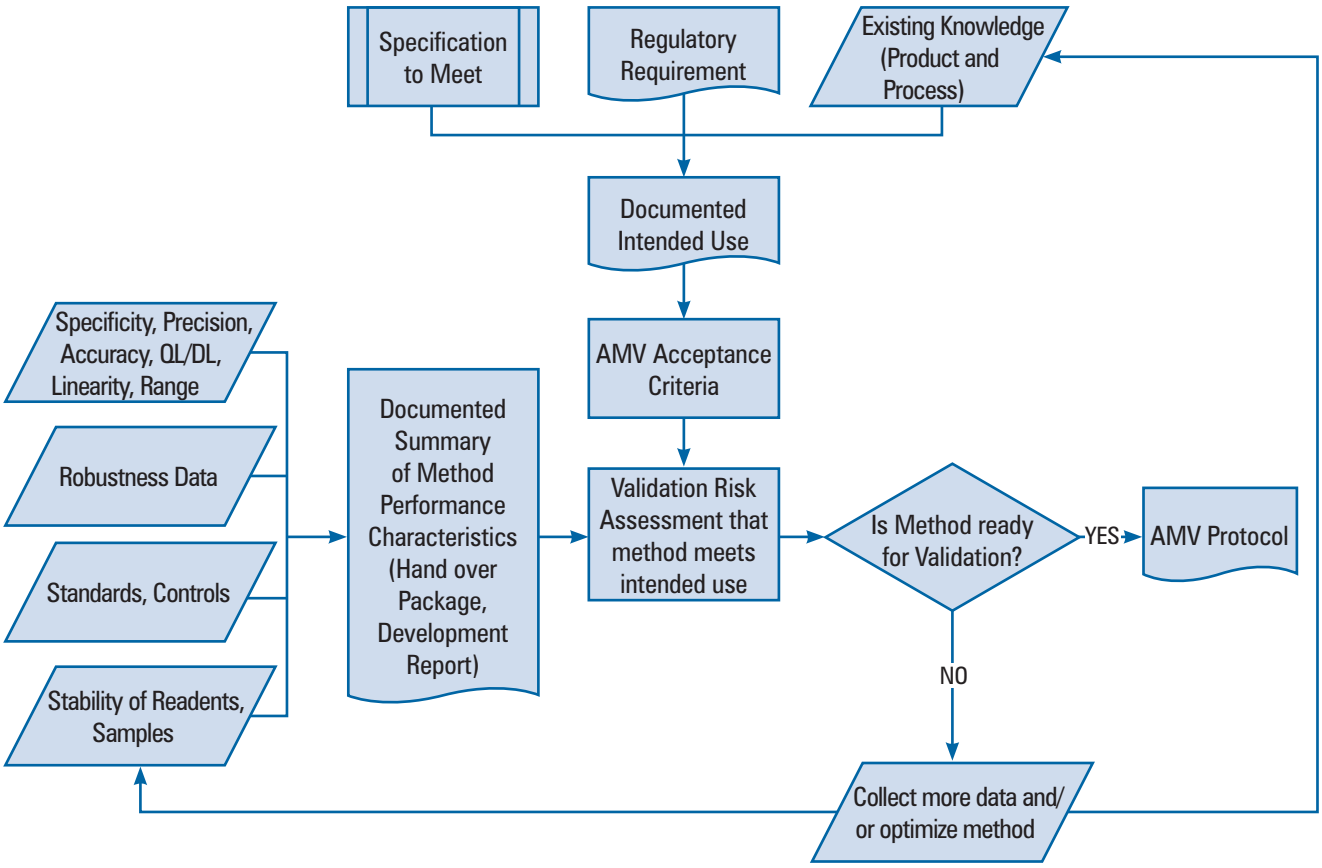
3.0 General Assessment of Method Validation Readiness

A risk-based assessment of the readiness of a developed/qualified method to be formally validated is an important part of AMV. A VMP could be used to describe the risk assessment process specifically for analytical methods. Once risks are integrated, suitable AMV study execution plans and protocol acceptance criteria can then be generated.

There can be significant time lags among AMD, AMQ, and AMV studies. Whether the AMD/AMQ to AMV transition is handled by the same person and/or department or occurs between different functional departments and/or locations, the validation readiness assessment followed should be similar.

The greater the understanding of the intended use of an analytical method, the risk involved to patient and firm, the expected test sample constitution, the production process, process capability, and the desired level of method performance, the easier it will be to assess validation readiness. This risk-based AMV process requires considerable effort evaluating data sources and quality expectations prior to AMV. A typical readiness assessment process is illustrated in **Figure 3.0-1.** Typical information sources, listed in **Table 3.0-1,** should be reviewed and assessed to determine the validation readiness.

Figure 3.0-1 Example of Assessment of Method Validation Readiness Flow Path



In **Figure 3.0-1,** the critical steps to consider when developing AMV acceptance criteria and the resulting validation risk assessment are shown as the three possible inputs in the top of this Figure. Not all inputs may apply for each AMV study, but should be considered whenever they do apply. As indicated in **Table 3.0-1,** the top three inputs constitute the intended use of the analytical method and lead to the development of suitable AMV protocol acceptance criteria.

The items on the left-hand side of **Figure 3.0-1** (historical method specific information) are used to assess the historical analytical method performance. Additional items, such as existing platform technology programs, may apply and are listed in **Table 3.0-1.** The specifications, historical process variation, and possible regulatory requirements dictate how widely AMV protocol acceptance criteria can be set depending on the risk(s) to patient and firm. The upper or maximum limits for the AMV protocol acceptance criteria should then be compared against the historical method performance, as summarized in the development report(s), qualification report(s), transfer report(s), assay control chart(s), and/or other available data sources. If the historical method performance is assessed to be significantly better than the maximum performance limits, the likelihood of failing the protocol acceptance criteria will then be low and the method can be considered ready for validation. In the event that the risk is deemed too high to proceed with AMV, at minimum, more data should be collected to provide more confidence that the method can perform as expected.

Table 3.0-1 General Method Readiness Assessment Guide

AMV Readiness Criteria (if available and/or applicable)	Data and/or Documents to be Assessed	Typical Points to Consider
1. Final test method procedure	1. Standard Operating Procedure 2. Method Change History Files	All material and processes need to accurately reflect the routine test method procedure. The method change history files should be reviewed to assess the potential impact of post-AMD and/or post-AMQ changes.
2. Intended use of method specified by:	VMP, method scope definition, (target) specifications	The product, production process step(s), and sample(s) to be tested should be well defined.
A. Specifications	2. (Target) specifications for raw material, in-process samples, release testing, and shelf-life	The number of significant digits used in the specifications, reflecting measurement precision capability, should be verified during AMV.
B. Regulatory and/or internal compliance expectations	3. ICH, local regulatory guidelines, current regulatory expectations, pharmacopoeia, internal procedures	AMV study designs may need to be adjusted to reflect current risk-based strategies and regulatory expectations.
C. Existing Knowledge (Product and Process)	4. Statistical process control data, scale-up studies, comparability studies, etc. (If applicable, similar product and process history records)	Actual process variation should be estimated from all known variation factors and considered for AMV protocol acceptance criteria.
3. Method performance capability judged by:	All relevant historical data sources should be reviewed	See below.
A. AMV characteristics (Accuracy, Precision, etc.)	1. AMD, AMQ, and AMT reports 2. Control charts	All existing reports and data should be reviewed to assess method readiness and capability.
B. Robustness	DOE results from AMD and AMQ studies	Robustness data should be used to limit the use of unsuitable analytical method process conditions.
C. Critical variation components	AMD and/or AMQ reports	Historical assay control data may also be used to estimate the impact of changes to method components (ex., analysts, instruments)

AMV Readiness Criteria (if available and/or applicable)	Data and/or Documents to be Assessed	Typical Points to Consider
D. System suitability criteria	1. AMD, AMQ, and AMT reports 2. Control charts 3. Invalid test records	System suitability criteria may not be included in test result reporting. Invalid test records can provide some additional indication of the robustness when system suitability criteria were set properly.
E. Reference standard	1. Reference qualification records 2. CoA 3. Stability records	The reference standard hierarchy should be considered ("gold standard", secondary standard, working standard). Stability should be assured.
F. Assay control	1. AMD, AMQ, and AMT reports 2. Control charts 3. Invalid test records	Assay control charting options should be considered. Valid assay control ranges should be adjusted to the desired method performance level.
G. Stability Reports/ Records	1. Stability Data	Existing stability data may provide a good assessment of long-term analytical method performance (Intermediate Precision).
H. Critical reagents and material	2. AMD, AMQ, and AMT reports 3. Intermediate Precision 4. Robustness	Reagents and material causing significant method result variation may need to be qualified and monitored. Alternate reagent and/or material sources may need to be evaluated if reagent/material supply is uncertain.
I. Analytical Instrument Qualification (AIQ) including software validation	1. AIQ report(s) 2. AMQ reports	All instruments used in AMV studies should be qualified. All software and/or spreadsheets should be validated. Analytical instrument capabilities and limitations should be well understood.
J. Reagent/Material (source) qualification	1. Vendor qualification reports 2. Material receiving documents 3. CoAs	Only approved vendors should be used for all reagents and material.
K. Analytical Platform Technology (APT)	AMV reports of comparable methods	When using APTs, historical performance data can often be used to set test system performance expectations for a comparable method.
4. Resource expectations	1. Personnel 2. Proficiency/Training requirements 3. Material and instrumentation supply	Sufficient personnel should be available to perform AMV studies. All personnel should be proficient and formally qualified for the functions performed. Sufficient material and instrumentation should be made available to execute AMV studies without significant interruptions.

Detail on assessing various risks and how to set suitable protocol acceptance criteria are provided separately in **Sections 3.1** and **3.2,** respectively.

3.1 General Risk Assessment Process

Several risk assessment tools can be used to facilitate risk-based AMV studies. For simplification, the risk assessment tools and processes are presented separately. Individual risk level results for each analytical method can be combined for each test method to be validated. The suggested risk assessment tools are examples; other acceptable alternatives exist. In summary, the goal of the systematic use of risk assessment tools and processes is to provide measurable results for:

- The desired amount of formal validation studies to be executed.
- The level of method performance needed as manifested in the AMV protocol acceptance criteria.

The amount of prospective AMV studies to be performed can be assessed by differentiating five general AMV situations, classsified as A–E in **Table 3.1-1, which** illustrates each of the five AMV situations, their typical risk and uncertainty level, and the resulting validation expectations ***(14).***

AMV class A–E are listed in order of the expected risk and/or uncertainty levels, with class A the highest expected risk and class E the lowest. Class A typically has the greatest uncertainty and risk to patient and/or firm because the relationship between product/process and the analytical method performance capabilities may not be fully captured. Thus, this higher level of uncertainty of suitability and performance requirements usually requires the generation of more data.

Previously validated and approved analytical methods and/or analytical platform technologies (AMV classes C and D), typically require less prospective AMV studies since the historical method performance is well captured. Class E (compendial methods) often requires the least amount of prospective (verification) studies as their use and suitability have been extensively demonstrated. However, prospective verification studies under actual conditions of use with representative test samples are still required for compendial methods.

The five different general AMV classes and resulting risk and/or uncertainty levels should be considered to be typical situations, however, the actual risk/uncertainly levels may be significantly higher for particular test methods. For example, the verification of the USP/EP Sterility Test (AMV class E) risk level may be higher than indicated here due to the severity of the potential impact on patients for false negative test results. Using another example, a higher risk/uncertainty level for a compendial method may also arise if a compendial potency method is used to replace an in-house potency method, because different potency results may affect the specifications and/or future dosing levels.

Table 3.1-1 The Five General AMV Classes and Prospective AMV Studies ***(14)***

AMV Class Description			Typical Risk / Uncertainty Level (1=Low, 5=High)	Suggested Prospective AMV Studies
AMV Class	Analytical Method	Product / Process Sample		
A	New	New	4–5	Full Validation
B	New	Old (Validated)	3–4[a]	Full Validation Plus AMC Studies
C	Analytical Platform Technology—minor change(s)[b]	New	2–3	Partial Validation
D	Old (Validated)	New	1–2	Partial Validation or Verification
E	Compendial	New	1–2	Verification per USP <1226>

a. If a new analytical method (forced method replacement) is needed due to supply reasons, the risk level can be generally considered higher because no other option may exist. Unforced test method replacements can be considered to be a lower risk level as more time may be available to optimize the method performance.

b. Some changes to validated APT methods such a different sample preparation step or the use of a different detection system may not require a full validation as only a part of the validated test system changes, whereas most of the system remains unchanged.

Further general risk assessments can be made from the consideration of analytical method types and their intended use. Analytical methods fall into one of the five general method type categories used to test for the identity, safety, purity, quality, and potency for biotechnological production processes. Risks can be assessed for analytical methods by considering the potential risks to patient if the analytical method fails to provide accurate and/or reliable test results. The following points to consider, as illustrated in **Table 3.1-2,** can be utilized to support the overall risk assessment associated with the use of a particular analytical method.

Table 3.1-2 Points to Consider in Overall Risk Assessment for Analytical Methods

Points to Consider	Example(s)	Expected Potential Risk/Impact
Method type and intended use (Identity, Safety, Purity, Quality, Potency, and Stability)	a. Safety Test: Sterility test using new rapid microbial method.	a. Potential risk to patients and firm is high if sterility test provides false negative results.
	b. Quality Test: Excipient concentration at final production stage.	b. Potential risk to patients is relatively low if the quality test provides inaccurate results as excipient is quantitatively added during production.
	c. Purity/Stability Test: Degradation products during storage.	c. Potential risk to patients is high if stability test is incapable to measure all degradation products.
Surrogate and/or complementary method is routinely used	Purity/Safety Test: A HPSEC method is used for quantitation of protein aggregate levels. A second electrophoresis method provides similar results for aggregate levels.	If second method routinely supports the results of the primary method, the risk to patients may be lower if the primary method provides inaccurate results.
Production Process Stage	Purity Test: Fermentation impurities are measured before purification and after purification.	Early-stage inaccurate impurity results from less reliable test method are lower risk to patients if late-stage testing provides more accurate results.
Sampling and Batch Uniformity	Potency Test: Potency testing in drug substance samples.	The potency results of in-process samples collected may be affected by the actual sampling process and/or hold times before testing. This risk may therefore be higher to firm as test results may not be representative of drug substance batch prior to filling.
Analytical Platform Technology (APT)	Purity Test: APT HPSEC method is used to test in-process samples.	Current QC experience with this method performance should lower the risk to patient and/or firm if the effect of different sample types is insignificant.

Once risks are understood and measurable, they should be controlled with suitable acceptance criteria in AMV protocols. Some additional potential risks to be considered are identified and defined in **Table 3.1-3.**

Table 3.1-3 General Risks to Patient and/or Firm *(14)*

Failure to meet acceptance criteria	"Wide" acceptance criteria
Risk to firm Potential inspection observations and overall compliance issues if failures are not completely resolved and justified before implementation.	***Risk to patient*** AMV results were near limits. This may lead potentially to unacceptable product because results may be inaccurate and/or unreliable.
Risk to firm Project progression/completion not possible or continued "at risk". Project completion could be significantly delayed and additional resources and time may be needed.	***Risk to firm*** OOS test results from inaccurate and/or unreliable test method may actually be within specifications and acceptable. Firm cannot release product that is actually acceptable.
Risk to patient Failed AMV studies may delay the supply of much-needed life-saving drugs.	***Risk to firm*** Any risk to patient is automatically also a risk to the firm.

3.2 Setting AMV Protocol Acceptance Criteria

3.2.1 Rationale

AMV protocol acceptance criteria should be set to balance two opposing considerations, thus preventing one of the two to dominate the other. The first consideration, to demonstrate a desirable high level of overall process and method capability within a given set of (target) specifications, may lead to setting narrow acceptance criteria for the analytical method performance. If the resulting method performance expectations, as manifested in the protocol acceptance criteria, are too narrow, it may become difficult to pass criteria during the formal validation execution studies.

The second consideration, to assure compliance and project completion by passing all protocol acceptance criteria, may therefore be directly opposing the first consideration. Acceptance criteria may be set unsuitably wide to assure that all criteria are readily passed. The method performance may therefore be considered validated, compliant, and acceptable although the actual method performance may not be suitable with respect to specifications, and/or overall process capability expectations. It is therefore important to recognize this relationship and set balanced acceptance criteria intended to satisfy both considerations as much as possible *(14)*.

When setting acceptance criteria for all relevant method performance expectations, it is essential to review all potential sources of variation and/or uncertainty that may impact the accuracy and reliability of test results and constitute a risk to patient and firm. Sources of variation/uncertainty should be reviewed, understood, and used to set acceptance criteria to ultimately ensure the suitability for use of the analytical method. The relationship of typical variation sources are expressed in **Equation 1.** For simplicity, the potential variation sources from the sampling process, transport, and storage, and/or the inconsistency in batch uniformity are considered to be part of the manufacturing process variation.

[Equation 1]

$$[\sigma \text{ mfg process observed}]^2 = [\sigma \text{ analytical method}]^2 + [\sigma \text{ mfg process actual}]^2$$

The (squared) observed manufacturing process variation is the sum of the (squared) variation sources of analytical method performance and the actual or true manufacturing process variation. As speci-

fications typically only exist for the observed manufacturing process variation, it is therefore critical to understand and control the underlying variation sources by using risk-based acceptance criteria for each of their maximum allowable variation.

3.2.2 Consistent Risk Assessment to Set Acceptance Criteria

Risk-based AMV protocol acceptance criteria should be predominately derived from the evaluation of two critical sources:

- (Target) specifications
- Existing knowledge — historical data of this product and/or process or similar products and process(es)

Other sources such as regulatory expectations may also impact acceptance criteria and should be considered when applicable. If the consistency of the sampling process and the batch uniformity is not an integral part of the manufacturing process variation or not known, these variation sources may also need to be considered.

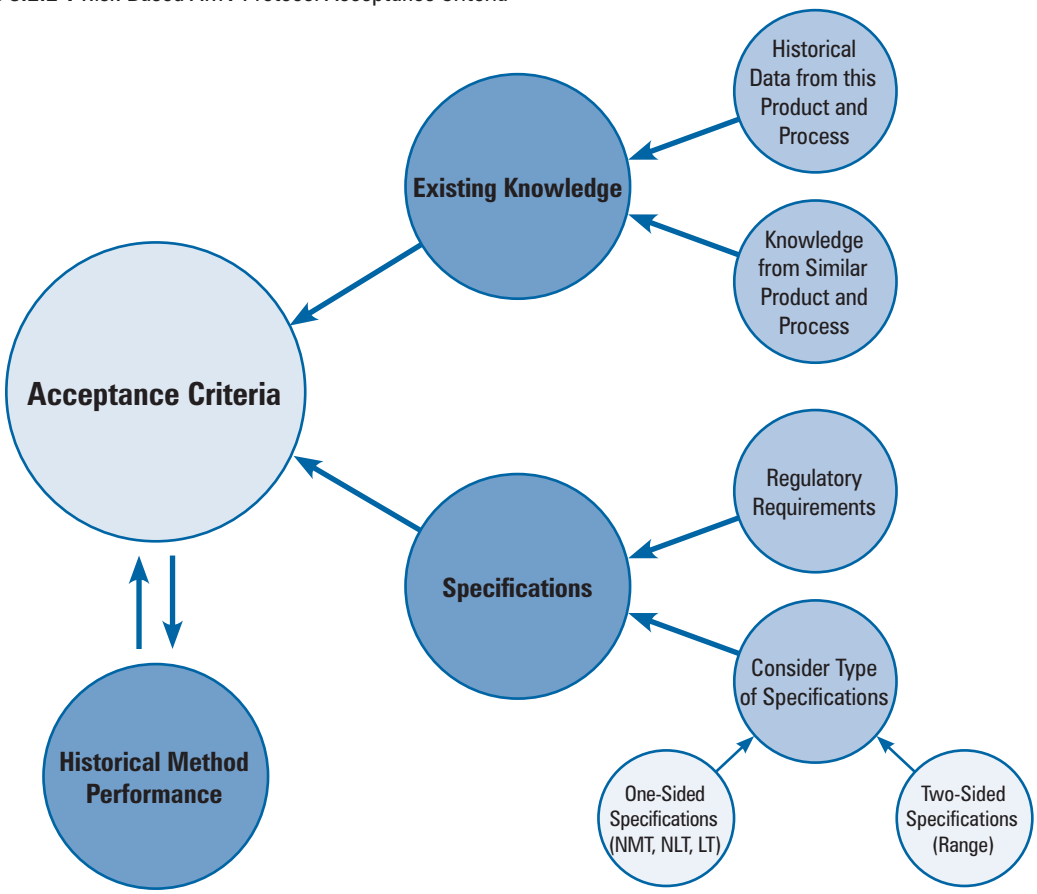
Figure 3.2.2-1 illustrates how AMV protocol acceptance criteria should be consistently derived ***(14)***. The specifications are the most extreme limits in which the total of all variation contributors should fall. In general, the higher the typical variation sources of production process and sampling process(es) are with respect to the specifications, the smaller the method variation needs to be for a method to be considered suitable for its intended use.

Acceptance criteria should be set to assure a minimum acceptable level of method performance. These method performance expectations are then compared to the existing historical data indicative of method performance capability. When starting this process by setting a certain desired performance requirement(s), these risk-based criteria should then be compared to the historical method performance data to provide an estimate on the likelihood of passing the acceptance criteria. Some balancing of the two opposing considerations may be necessary. However, if the historical method performance data sources do not provide sufficient evidence, or the method is simply not capable, then the method may not be ready to proceed to AMV studies.

Some AMV protocol acceptance criteria, such as linearity regression coefficient(s), may not be directly connected to visible method and/or process capability indicators. For example, the linearity regression coefficients typically used in system suitability and/or AMV acceptance criteria describe the best-curve fit over an expanded routine assay range. Limits set for the overall curve fit, although useful for controlling routine testing, cannot be directly compared to overall accuracy and/precision performance requirements. In those cases, acceptance criteria could be set from the historical system suitability data.

In those cases where only a few data points exist to estimate true process and/or method capabilities, other reference points, such as previously approved APT method performance criteria, can be used to assist in setting acceptance criteria. Regardless of which references are used to set acceptance criteria, the method performance limits set should assure that the method will produce accurate and reliable results well within the specifications.

Figure 3.2.2-1 Risk-Based AMV Protocol Acceptance Criteria



3.3 Example for AMV Protocol Acceptance Criteria

A new container closure system for a previously validated manufacturing process (including AMV studies) is leading to a partial revalidation of a final drug product test. The analytical method itself is unchanged and the specifications have remained the same (AMV class D, see **Table 3.1-1).** Additional studies and risk assessments provided sufficient evidence that the impact on release and stability testing is not expected to be significant and a full validation study may therefore not be required. A partial, formal, prospective AMV study for the test is set up to verify the method performance characteristics, accuracy, and intermediate precision are still within acceptable limits for the new container closure system. The lack of interference from the new container closure system (specificity) is inferred from the accuracy results. **Table 3.3-1** lists the historical data for this AMV category D validation (modified process/product) that should be reviewed to set AMV protocol acceptance criteria.

Table 3.3-1 Historical Data for Manufacturing Process, Assay Performance, and Suggested Limits for Accuracy and (Intermediate) Precision

Specifications (using old and new container closure systems)	***90 – 110 %***	
Statistics	**Mean (in %)**	**SD (in %)**
Statistical Process Control: Manufacturing Process Performance (last n=30)	100.2	5.0
Assay Control Performance (last n=30)[a]	102.3	***3.0***
From Previous AMV Studies: Intermediate Precision of Assay Control (total n=36)	99	***2.0***
From **Equation 3.2.1-1:** Actual (True) Process Performance (estimate)	***(100)***	***(4.0)***
Suggested AMV Limits for Overall Accuracy	98 – 102	—
Suggested AMV Limits for Intermediate Precision	—	3.0

a. The current assay control limits were set during/after AMV studies and have remained unchanged. Results are routinely reported to the specification units (100 %) while descriptive statistics are given to the 1/10th of the reported unit (100.0 %).

3.3.1 Setting and Justifying Acceptance Criteria for the AMV Protocol

The specification range of 90 – 110% is used as the primary reference range. The total variation of all contributing processes should fit well within this range. **Table 3.3-1** lists the relevant specifications and the observed historical process/product and method performance data (bolded numbers). The true or actual process variation (4.0 %) and the historical method performance, as judged by the assay control variation (3.0 %), are estimated as follows:

$$[5.0\%]^2 = [3.0\%]^2 + [4.0\%]^2$$

The actual process variation is approximately 4.0%. When using the suggested Accuracy acceptance criteria (98 – 102 %), a resulting worse-case situation of a process target or midpoint shift of about 2% is considered acceptable with respect to the impact to patient and firm. The specifications are still at least 2 standard deviations away from the future worse-case (acceptable) bias of 2%. The previous AMV study results for accuracy of 99 % mean recovery, in addition to the recent assay control data (unchanged), suggest that the suggested AMV acceptance criteria should be readily passed and appears to be sufficiently balanced.

When using the suggested protocol acceptance criteria for Intermediate Precision (3.0 %) under similar AMV conditions as previously executed (2.0 %; n=36 total results for Intermediate Precision), an equal-or-better performance should be likely when compared to the historical assay control performance (3.0%) which was generated over several months. The assay control data was generated with a maximum variety of method components over an extended period of time, thus providing a worse-case long-term method variation estimate. Compared to the previous AMV study results (2.0 %) and the historical assay control data variation (3.0%), the acceptance criteria for Intermediate Precision appear to be properly balanced.

4.0 Analytical Method Validation

ICH Q2(R1) provides basic guidance for AMV studies to assure product safety, efficacy, and quality *(1)*. The practical guidance provided in this TR builds on ICH Q2(R1) and the relevant risks illustrated in **Section 3.0.** Once analytical methods are considered suitable and ready for AMV studies, guidance provided in this section can be used to prepare AMV protocols and reports.

Table 4.0-1 summarizes the recommended validation parameters for each type of test procedure and intended use per ICH Q2(R1) guidance. It is important to recognize that the suggested ICH Q2(R1) validation characteristics are minimum expectations and that more studies and evidence may have to be considered based on risks identified in **Section 3.**

Table 4.0-1 Minimum AMV Characteristics Per ICH Q2(R1)

ICH Q2(R1) Category (Test)	I (Identification Test)	II (Quantitation of Impurities)	III (Qualitative Limit Test for Impurities)	IV (Quantitation of Active Ingredient)
Accuracy	No	Yes	No	Yes
Repeatability Precision	No	Yes	No	Yes
Intermediate Precision	No	Yes[a]	No	Yes[a]
Specificity[b]	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[c]	Yes	No	No

"Yes" signifies that this characteristic is normally evaluated.

"No" signifies that this characteristic is normally not evaluated.

a. In cases where reproducibility has been performed, intermediate precision is not needed.

b. Lack of specificity of one analytical procedure could be compensated by other supporting procedure(s).

c. May be needed in some cases.

Table 4.0-2 summarizes typical minimum AMV study execution plans, all relevant reported results, and the acceptance criteria to be developed and justified. The result of DOE studies performed prior to AMV studies to support robustness, may not have to be repeated in the AMV studies. The number of replicates and/or concentration levels should be adapted to the intended use of the analytical method and all identified risks. Analytical methods with relatively high levels of variation in the reported results may require more replicates and/or concentration levels as the minimum requirements listed in **Table 4.0-2.**

Table 4.0-2 ICH Q2(R1) Requirements and Suggested Reported Results and Acceptance Criteria ***(14)***

ICH Q2(R1) Characteristic	Minimum Requirements Per ICH Q2(R1)	Typical AMV Results Reported	Acceptance Criteria Should be Developed for the Following Results
Accuracy	Minimum of 9 determinations over the specified range (3 concentrations/3 replications each).	Mean % recoveries for each concentration; Overall % recovery.	Mean % recoveries for each concentration; Overall % recovery.
Repeatability Precision	Minimum of 9 determinations over the specified range (3 conc./3 repli. each) and a minimum of 6 determinations at 100% of test conc.	Mean(s), standard deviation(s), CV(s), appropriate number of significant digits to be reported. Confidence Interval(s) for SD(s).	Standard deviation(s) and/or CV(s).
Intermediate Precision	Use an Intermediate Precision Matrix, 3 levels for each factor is recommended. Two levels are a required minimum. Data should be analyzed by ANOVA. Other and/or additional approaches should be justified.	Factor standard deviations and CVs, factor means, Individual and overall p-values of factors associated with ANOVA, overall CV, difference between most extreme factor means. Confidence Interval(s) for SD(s).	Overall CV, Individual and overall p-values of factors associated with ANOVA, Difference between most extreme factor means for $p<0.05$).
Reproducibility Precision	Needed when different locations will perform testing. Minimum of 3 determinations for each factor for each laboratory.	Factor standard deviations and CVs, factor means, Individual and overall p-values of factors associated with ANOVA, overall CV, difference between most extreme factor means. Confidence Interval(s) for SD(s).	Overall CV, Individual and overall p-values of factors associated with ANOVA, Difference between most extreme factor means for cases when $p<0.05$.
Specificity	Provide evidence that analyte and matrix interferences are negligible.	P-value of difference testing, means, standard deviations and/or CVs, difference between most extreme factor means.	p-value for difference testing, Difference between most extreme factor means for $p<0.05$).
Detection Limit	Base on visual approach or signal-to-noise (3:1 or 2:1) or standard deviation of response and slope or other justified approach.	Detection Limit, calculations or graphical representation of Detection Limit.	Detection limit.
Quantitation Limit	Base on visual approach or signal-to-noise (10:1) or standard deviation of response and slope or other justified approach.	Quantitation Limit, calculations or graphical representation of Quantitation Limit.	Quantitation limit.
Linearity	Minimum of 5 concentrations; Justify other approaches.	Scatter plot of individual data points, r (or r^2), y-intercept, slope, residual (error) sum of squares.	Minimum r or (r^2) value.
Range	Consider these minimum ranges: **Active ingredient:** 80-120% of test conc. **Impurities:** reporting level to 120% of upper specification limit.	Range is the assay range over which the assay provides acceptable Linearity, Accuracy, and Precision.	Minimum required Range. Within the Range, all acceptance criteria for Linearity, Accuracy, and Precision should be passed.

4.1 AMV Characteristics

ICH Q2(R1) AMV characteristics are briefly described, along with points to consider for ensuring successful AMV studies. Additional considerations may apply based on the method and/or sample types, the intended use of the method, the available historical data, and other reasons.

4.1.1 Accuracy

Accuracy is most often confirmed by spiking a reference material into the product matrix and calculating percent (net) recoveries over the desired assay range. Dilutions with matrix buffer could be done to cover the lower part of the assay range.

Values for the expected readings to calculate percent recoveries may be difficult to calculate because the reference standard used may not be 100% pure and/or may not yield an identical response signal versus the analyte of interest. All possible imperfections for the quantitative preparation of spike and dilutions samples should be controlled.

Several strategies can be used to confirm Accuracy during the AMV studies. Although the strategies to demonstrate Accuracy can greatly vary depending on the method types, available reference material, and potential risks, a general ranking order of most-to-least preferred strategies, 1-6, is provided below. Additional options may also exist. If purified reference material is not available and/or an absolute reference value is not known, accuracy may be inferred once precision, linearity and specificity have been established.

1. Spiking of suitable reference material into sample.
2. Spike concentrated product (and/or impurities, etc.) and/or dilute to cover entire range.
3. Determine Accuracy by direct comparison of a recognized (reference) method.
4. Infer Accuracy only from dilution and/or concentration studies.
5. Analyze only reference or calibration material independent of product and/or matrix.
6. Compare results only to an alternative method which is not recognized.

Accuracy of the reported results should be demonstrated with respect to the reported units. For example, when specifications and corresponding analytical results for impurities are reported in percentages of the total of all (potential) product components in the sample, the recovery calculations are typically expected to be expressed as concentration levels.

To demonstrate acceptable recoveries, t-statistics can be used to compare means of observed vs. expected percent recoveries. The results of the AMV studies for Accuracy can be directly affected by how accurate and reliable reference standard solutions and spiked samples are prepared and handled. Data generated for Accuracy may be used to cover required data for other validation characteristics, such as Repeatability Precision, Linearity, Range, and, if needed, the Quantitation Limit.

4.1.2 Repeatability Precision

Repeatability Precision indicates how precise the results are under best-possible testing conditions. Repeatability Precision should be demonstrated over or beyond the entire assay range.

To demonstrate acceptable Repeatability Precision, the (percent) coefficient of variation (%CV) values at different concentrations can be compared. The demonstration of Repeatability Precision can be potentially affected by how well random errors in sample preparation can be controlled.

4.1.3 Intermediate Precision

Intermediate Precision results reflect the expected precision of a test method in day-to-day laboratory operations. Typical day-to-day changes should be simulated in the AMV studies to evaluate this validation characteristic. Intermediate Precision should be demonstrated by using typical representative sample(s). Depending on the intended use of the analytical methods and the identified risk(s), additional samples representing the expected assay range could be integrated into the Intermediate Precision studies. This data should ideally be generated by sets of 3, although this may not always be possible (for example, only 2 instruments are available).

Results can be generated as a partial factorial design by rotating operators, days, instruments, and possibly other critical factors identified during the AMD, AMQ and/or historical testing. This relatively simple execution matrix design is illustrated in **Table 4.1.3-1**. An ANOVA, where results can be grouped by each operator, day, and instrument and analyzed in one large table, can be used as the statistical tool.

Table 4.1.3-1 Intermediate Precision Matrix

Sample	Day	Operator	Instrument
3x	1	1	1
3x	1	2	2
3x	1	3	3
3x	2	1	2
3x	2	2	3
3x	2	3	1
3x	3	1	3
3x	3	2	1
3x	3	3	2

A mixed-linear model analysis can also be used to assess the Intermediate Precision results ***(15).*** A numerical secondary limit may be used to control the likelihood of observing statistical differences due to high precision and some differences (bias) are normal and should be expected. The mixed linear model analysis can be more practical and useful than ANOVAs. An example for the mixed linear model analysis results for an automated ELISA method is shown in **Table 4.1.3-2** (using the execution matrix of **Table 4.1.3-1**) ***(14).***

When interpreting intermediate precision results in this example by the mixed-linear model, a significant amount of variation is observed among the three different instruments used. This should be considered when changing or qualifying additional instruments. The unidentified residual variation (CV = 11.6%) may include other contributing variation sources such as random sample handling variation during the AMV studies and the variation represented by Repeatability Precision within the triplicate sets used for each of the variation source combinations. The unidentified residual variation (CV = 11.6%) could be compared to the AMV Repeatability Precision results and further evaluated, if needed, with the intent to improve the overall test method's intermediate precision (CV = 14.6%).

Table 4.1.3-2 Mixed Linear Model Results for Intermediate Precision Matrix ***(14)***

Effect	Variance	Std Dev.	CV
Overall	0.0249	0.158	14.6%
Instrument	0.0158	0.126	11.6%
Operator	0.0013	0.036	3.3%
Day	0.0004	0.020	1.9%
Residual	0.0157	0.125	11.6%

4.1.4 Reproducibility (Precision)

As an alternative to the AMT strategies illustrated in **Section 5,** Reproducibility of test results among multiple laboratories is the last of the three precision classes that can be established as part of the AMV studies. ICH Q2(R1) defines Reproducibility as an inter-laboratory trial with the intent to standardize an analytical procedure. ICH Q2(R1)'s Reproducibility studies are not intended to be part of the marketing authorization dossier. Reproducibility is therefore not covered here as part of AMV studies, however, the comparison of analytical method performance between multiple laboratories is addressed in **Section 5.**

4.1.5 Specificity

Specificity of an assay is usually ensured by demonstrating, **a**) no or only insignificant matrix interference, and, **b**) no or only insignificant interference from other potential analytes that could be present in the matrix. Because the sample matrices of biopharmaceuticals are usually complex and can vary, potential interferences should also be evaluated by spiking other analytes at worse-case concentration levels.

A DOE set up of all routine and potential analyte and/or matrix components at their relevant specification or expectation levels should provide valuable information for Specificity.

To assess matrix and and/or analyte interference, spiked samples can be compared to unspiked samples using comparative statistics. 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. A secondary numerical limit for the allowed maximum difference(s) can be established, in case $p < 0.05$, which could be similar to the limit stated under the validation parameter Repeatability Precision.

4.1.6 Linearity

The Linearity AMV studies demonstrate proportionality of signal response and results versus analyte concentrations over the intended assay range. Linearity is evaluated through a linear regression analysis which uses individual results of either analyte concentration versus assay results, or observed versus expected results. Similar to Accuracy studies, all samples prepared for Linearity should contain the desired levels of the actual analyte of interest within a typical sample matrix.

For observed versus expected analyte concentration regression slopes significantly different from 1.00, possible reasons should be assessed and can often be confirmed from the results of the other validation characteristics. For example, differences could arise from an inhibitor effect from matrix components. Y-intercepts significantly greater or less than 0 with a slope of 1.00 could result from sample preparation errors or other similar reasons. A correlation coefficient (r) significantly less than 1.00 may arise from various reasons, for example, lack of precision, accuracy, narrow assay range, or

poor sample preparation, and should be confirmed with the results of other validation characteristics whenever possible.

ICH Q2(1R) requires reporting the regression line y-intercept, slope, correlation coefficient, and the residual sum of squares. Acceptance criteria for the y-intercept, slope and the correlation coefficient should be derived and justified for Linearity. The method type and intended use can be used to set acceptance criteria. The observed statistical differences (e.g., 90% confidence intervals) will be compared with the acceptance criteria to indicate overall system suitability.

A pseudo-linearity may exist for many assays even after mathematical transformations. Whenever a pseudo-linearity analytical method response curve exists, the required assay range should be considered when the assay response curves are "forced" into linearity, as the resulting validated assay range may be too narrow for routine testing. The potential limitations of forced or converted analytical method response curve linearity should be considered prior to the AMV studies.

4.1.7 Range

The Range of an analytical test method must bracket the specification range. By definition, the QL constitutes the lowest point of the Range and is the lowest analyte concentration that can be quantitated with acceptable accuracy and precision. In some cases, the QL may also apply for the upper part of the assay range. In addition to the required Accuracy and Precision for all analyte concentration points within the Range, the analytical method response curve may be linear or pseudo-linear and this should be considered as stated in Linearity.

AMV results for Range may be evaluated from the same data set(s) as Linearity and/or Accuracy. AMV protocol acceptance criteria for Range should therefore be similar to those of Accuracy, Repeatability Precision, and Linearity. The potential limitations of the resulting restrictions to the suitable Range should be considered.

4.1.8 Detection Limit (DL)

The DL for a specific analyte in a given sample matrix can be described as the analyte concentration giving a signal significantly different from the blank or background signal. This (positive) signal will be observed almost every time the analyte is at or above the DL concentration. It is important to consider for the overall control strategy that there is often an acceptable residual probability that an analyte at the DL will not be detected. For example, using the visual signal-to-noise ratio of 2:1 in a chromatographic method, the relatively high probability of not detecting the analyte may have to be counterbalanced by multiple testing. Whenever the risks associated with undetected analytes at or above the DL are significant, multiple testing for routine samples should be considered to compensate for this potential performance gap.

ICH Q2(1R) suggests three different approaches to determine the DL. Other approaches may also be acceptable when these can be justified. Following ICH Q2(1R), the DL may be determined by visual inspection (A), signal-to-noise ratio (B), or the standard deviation (SD) of the response and the slope (C).

For approach A, B, or C, and any other justified approaches, the DL should be significantly lower than the desired specifications. A particular approach should be justified with a clear expectation of method capabilities and the intended use.

4.1.9 Quantitation Limit (QL)

The QL is by definition the lowest (and highest) analyte concentration that can be quantitated with

accuracy and precision. Since the QL constitutes the beginning of the Range of quantitative limit tests the Range criteria for Linearity should be passed for the low analyte concentration determined to be the QL. The determination of the QL may involve similar ICH Q2(R1) approaches (A, B, and C) as those used for DL.

When identifying the ICH Q2(R1) approach and/or any other suitable approaches in the AMV protocols, several points should be considered. Multiple replicates of analyte concentrations tested (instead of averages) yield typically higher standard deviations and therefore higher QLs. Approach C may provide low QLs when the assay response is highly linear, precise, and accurate over the selected Range. Spiked sample preparations should be accurate and precise to prevent random and systematic deviations from the regression line as they may increase the QL. Low analyte concentrations are preferred for use to determine to QL because high concentration outliers may disproportionally increase the QL due to an increased "leverage effect".

4.1.10 Typical AMV Execution Matrix

A typical AMV execution matrix for a quantitative limit test is illustrated in **Table 4.1.10-1** ***(14).*** The AMV execution matrix should include the minimum validation characteristics required per ICH Q2(R1). Data generated for accuracy may also be used to cover required data for other validation characteristics, such as repeatability precision, linearity, assay range, and QL. This will reduce the amount of data generated and the likelihood of random errors based on multiple sample preparations. Unexpected results obtained for a particular validation characteristic should also be observed for other validation characteristics using the same data set.

Table 4.1.10-1 Typical AMV Execution Matrix for a Quantitative Limit Test

ICH Q2(R1) Validation Characteristic	Analyst Number	Day Number	Instrument Number	Validation Methodology (Spiked Analyte Concentration)
Accuracy	1	1	1	(3 x each): 0.5, 1, 2, 5, 10, 20, 40, 60, 80, 100, 120%
Repeatability	1	1	1	Same as accuracy; plus n = 6 repeats at 100% concentration
Intermediate Precision	1	2	1	3 x 100% concentration
Intermediate Precision	2	2	2	3 x 100% concentration
Intermediate Precision	3	2	3	3 x 100% concentration
Intermediate Precision	1	3	2	3 x 100% concentration
Intermediate Precision	2	3	3	3 x 100% concentration
Intermediate Precision	3	3	1	3 x 100% concentration
Intermediate Precision	1	4	3	3 x 100% concentration
Intermediate Precision	2	4	1	3 x 100% concentration
Intermediate Precision	3	4	2	3 x 100% concentration
Specificity	1	5	1	Matrix interference
Specificity	1	5	1	Analyte interference
Linearity	1	1	1	Same as accuracy
Assay Range	1	1	1	Same as accuracy
QL	1	1	1	Same as accuracy (but only 0.5 to 10% range)

4.2 Additional AMV Characteristics to be Considered

The following method performance criteria, listed in **Table 4.2-1,** are ideally captured during AMD and/or AMQ studies but should be repeated and/or summarized in the AMV protocol or report. For any prospective studies conducted during the AMV studies, **Table 4.2-1** also includes the suggested reported results as well as the corresponding prospective protocol acceptance criteria.

Table 4.2-1 ICH Q2(R1) Requirements and Suggested Reported Results and Acceptance Criteria ***(14)***

Analytical Method Performance Characteristic	Retrospective (AMD/AMQ) or Prospective Evaluation During AMV Studies	Report the Following Results	For Prospective AMV Studies, Acceptance Criteria Should be Developed for the Following Results
Robustness	Deliberately perform minor changes to critical assay parameters such as incubation temperature or time DOE matrix	P-value associated with statistical analysis, Means, standard deviations and/or CVs (if replicates), difference between most extreme factor Means	P-value associated with statistical analysis, discuss the importance of p-values <0.05, difference between most extreme factor means for $p<0.05$). The method should be robust.
Signal Response Factors	Establish analyte response factors whenever multiple components are present	Response factors; if necessary, normalization values	No significant differences among response factors should be observed. If significant differences are observed, normalization factors should eliminate these differences.
Statistical Data Reduction	Establish analyte response curve statistics (e.g., linear regression)	Slope(s), y-intercept(s), regression coefficients, suitable range, sample calculation(s)	Software/spreadsheets should be validated. The number of generated invalid tests should be small.
Degradation (for Stability-Indicating Methods)	Establish stability profile and degradation pathways of samples, impurities, and by-products	Identify degradation pathway and products. Report percent degradation for each product storage or handling condition (e.g., freeze/thaw, heat, dilution, etc.)	Some degradation or other changes to test results should be visible for all stability-indicating test methods.
Stability of All Material	Evaluate the short-term (during testing) and long-term (during storage) stability for samples, standards, controls, reagents, and material.	Short-term and long-term storage conditions (temperature, time, and container)	All material should be stable during the permissible testing time (per SOP). All material should be stable over the assigned expiry time (per SOP).
System Suitability	Establish that components of the test system are suitable for routine testing	Conditions or criteria to be met to ensure that the test system is controlled and provides valid data	Specify requirements and method(s) of verifying conformance to requirements.
Sample Suitability	Establish that sample and/or testing replicates are appropriate to routinely support accurate and reliable test results	Conditions or criteria to be met to ensure that sample suitability is controlled and provides valid data	Sample size (replicates) and resulting precision limits. Sample size could be modified based upon the AMV data evaluation.

Analytical Method Performance Characteristic	Retrospective (AMD/ AMQ) or Prospective Evaluation During AMV Studies	Report the Following Results	For Prospective AMV Studies, Acceptance Criteria Should be Developed for the Following Results
Significant Digits	Generate data under Repeatability Precision conditions	Results per ASTM E29-02 or other recognized procedures ***(16)***	Significant digits in reported results should be equal or more than corresponding significant digits in the specification(s).
Analytical Method Comparability (see **Section 6)**	Establish the mean difference of control results for new versus old method; modify in-process and/or product specifications if necessary based on a sufficiently large number of data pairs	T-test statistics	Overall CV for each method. P-value from t-test. A significant difference between both method means may lead to specification adjustment(s).

4.2.1 Assay Bias and Analytical Response Factors

The evaluation of Accuracy within complex biological matrices can be challenging when gold standards do not exist. It is therefore important to establish analytical response factors as early and accurate as possible for methods which measure relative percentages of various (potential) analytes. Many quantitative specifications are based on concentration levels and analyte response factors should therefore be directly related to concentration levels. Significant differences in response factors should be integrated (normalized) in the calculations to permit the reporting of accurate analyte levels.

4.2.2 Stability of Samples, Standards, Controls, Reagents, and Material

Samples, standards (secondary, in-house, or working), controls (external or in-house), and all critical reagents and material should be evaluated for degradation during storage and potential freeze-thaw cycles. The maximum limits for bench time during actual testing (room temperature), repetitive freeze-thaw cycles, and long-term storage of all materials used to generate test results should be evaluated and storage and expiration times should be established based on actual data. Whenever internal and/or external laboratory data is not available, some prospective data may have to be generated to establish or confirm storage and handling conditions. **Table 4.2.2-1** provides an example for a prospective test plan for a critical reagent's expiry time.

Table 4.2.2-1 Prospective Expiry Date Study Protocol for a Critical In-House Reagent *(14)*

Introduction and Background	Expiry Scope and Strategy	Acceptance Criteria
The current expiry date for this critical in-house reagents of one month will be assessed from prospective study data. This reagent is routinely prepared by qualified analysts under controlled conditions and stored in controlled temperature environments (2 to 8 °C) when not in use. The average exposure time to room temperature (20 to 25 °C) is not more than 3 hours.	To confirm the 1-month expiry date for this critical reagent for the testing under this SOP, several time points will be evaluated for assay performance using the aged critical reagent. This prospective evaluation will start with testing of the assay control using fresh preparations of this critical reagent (at time zero) and continue with testing the assay control at specific time points during the proposed expiry date period. The time points will be evaluated statistically for differences in mean results. For the one month outdates for the reagent, measure (n=6) at t=0 (first week) t=2 weeks, t=3 weeks, and t=4 weeks. If possible, the same instrument and operator will be used for each test point. In addition to the assay control data, the pH will be measured and a visual inspection will be performed for each time point. For the expiry date evaluation, the longest time period(s) that starts at the day of the reagent preparation and ends at the last day of testing (data not statistically different from time t=0), will be used as the relevant expiry date.	For a time point to be suitable for the outdate assignment, all system suitability criteria should be met as indicated by an acceptable t-test result ($p>0.05$) when testing for the difference in mean results. If $p<0.05$, the difference of means between the particular time point and t=0 should be NMT 2SD from the recent historical control data trend chart mean.

4.2.3 System Suitability

The test system suitability should be properly controlled to ensure reliable test results. The system suitability criteria should be built up during AMD and/or AMQ, and finalized during or after the optimization phase. This is usually accomplished by running a set of control points, such as the performance of the assay control or calibration curve (ex., regression coefficients, curve reproducibility limits). Later during validation and routine testing, when system suitability will be passed (all control points are within established limits), therefore providing valid and suitable assay run conditions.

For many assays, the validity of the daily performance of the entire test system which includes operators, test reagents and materials, all equipment and instrumentation, software and spreadsheet calculations, standards, controls, samples, and the statistical data reduction is essentially judged by the assay control value(s). Basic statistical process control (SPC) principles are commonly used here, that is, any assay control result outside of established limits (for example, ± 3 standard deviations (SD)) is considered an outlier, thus resulting in an invalid assay run that should be repeated to obtain valid test results.

4.2.4 Sample Suitability

Sample suitability can also be evaluated as part of system suitability. Sample suitability should be established during AMD and/or AMQ, and should ideally ensure that samples, controls, and standards are prepared identically as much as possible (same dilutions and final matrix, etc.) and run simultaneously within the same assay run. This should ensure that these important test system components are locked in and captured in the (draft) SOP before AMV is started. This should also ensure that reference standards are providing accurate results for samples and controls, and the performance of the controls can be considered indicative of the test system performance over time. This is important later when extending the validation status (ex., new operator, instrument, reference standard, testing laboratory, etc.) and criteria for suitability are derived from the assay control performance.

Sample suitability should include a statistical analysis of the number of replicates that should be run to generate release results. For example, if the production process sampling can deliver true batch-representative samples, and the assay repeatability precision is very high (small variation) when compared to the product specifications (and therefore the high degree of batch-to-batch variation that these specification are based upon), then single measurements may be acceptable. Often however, assay precision for the testing of biopharmaceuticals is relatively low and multiple measurements will significantly increase the level of certainty in the corresponding test results. The sampling of batches in itself may have significant variability. Several detailed and recognized standard practices are available ***(17,18).***

4.2.5 Statistical Data Reduction

Statistical results describing the overall data fit for an assay run, compared to historical assay runs (for example, regression line correlation coefficients) should be used to confirm that a test system performance is acceptable.

Typically, fifteen assay runs are considered to be the minimum data set to compare statistical models such as 4-parameter and 5-parameter logistic curve functions, with and without weighing factors. Just as different test methodologies have bias, changing statistical models may significantly change the final results generated. Some models may simply be inappropriate as they do not provide acceptable results over the desired assay range.

4.2.6 Robustness

Small but deliberate changes to the test system should be evaluated during AMD and/or AMQ once the method is considered to be optimized. Changing operational limits such as reaction times or temperatures at or after the AMV studies may invalidate the AMQ and/or AMV results if these operational limits significantly affect the test system performance.

Analytical method elements which are critical, meaning they significantly contribute to overall test result variation, and therefore, result in insufficiently optimized test method performance, need to be understood and controlled through operational SOP limits during the later stages of the optimization phase, before system suitability criteria are finalized. The result of DOE studies performed prior to the formal AMV studies, for the confirmation of suitable Robustness of the analytical method, may not have to be repeated in the AMV studies if no significant change(s) to the analytical procedure occurred after the DOE studies.

Several recognized documents from ASTM Guides provide guidance on how to set up DOEs, interpret data and reduce test method variability ***(17-20).*** It should be noted, however, that ASTM uses the word 'ruggedness' to describe what is commonly considered (ICH Q2(R1)) as 'robustness', whereas the United States Pharmacopoeia (USP) 34, <1225>, defines 'ruggedness' as varying test method variance components ***(21).*** Varying test method components and the expected variability in test results is commonly considered (ICH Q2(R1)) as Intermediate Precision.

4.2.7 Degradation

Degradation of test materials is usually evaluated and limited as part of the test material stability studies. The stability of the assay control and sample can be assessed and controlled as part of sample suitability. Whenever the intention is to use this method as a stability-indicating method (usually for drug substance and drug product) as part of the stability program, forced degradation studies should be performed.

Stability-indicating test methods that are used to assess the stability of the drug product and which are

used to assign and monitor the shelf-life of the drug product under less degrading storage condition, (ex., stored at 5 degrees Celsius), should indeed be capable to measure degradation or changes to the drug product. Two ICH guidance documents (ICH Q1A and Q5C) provide a good basic overview of how the stability testing program is to be set for analytical methods and how data generated should be used to provide evidence for the desired shelf-life ***(22,23).***

4.2.8 Significant Digits in Reported Results

Because the uncertainty in test results for biopharmaceutical products is usually not expressed by using confidence limits, the uncertainty in results should be expressed by using the appropriate number of significant digits. The appropriate number of significant digits to be used for reported results should be based on the assay precision and could be determined during the AMV studies.

A simple way of demonstrating the proper use of significant digits is to use a widely accepted standard reference procedure such as ASTM E 29-02 ***(16),*** which provides clear instructions for generating significant digits from repeatability precision, as required of quantitative AMVs per ICH Q2(R1) guidelines. The reason that AMV studies should deliver the appropriate reported uncertainty for test results lies mostly in the fact that by the time an AMV is executed, a final version of the test method SOP is available and QC operators have been trained ***(24).***

Following this ASTM E 29-02 practice, in which the definition for repeatability precision matches the definitions used in ICH, USP and FDA guidelines, provides the advantage of having a reference to an accepted consensus standards organization document. ASTM's process to generate the number of significant digits in reported results also allows the user to retain more significant digits that other approaches may allow. 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 (s). The applicable standard deviation is the repeatability standard deviation (see Terminology in ASTM E 456) ***(25).*** Test results should be rounded to not greater than 0.5 s or not less than 0.05s, provided that this value is not greater than the unit specified in the specification (see **Section 6.2**). When only an estimate, s, is available for s, s, may be used in place of s in the preceding sentence" ***(16,25).***

An example how to derive the appropriate number of significant digits from AMV Repeatability Precision is given in **Table 4.2.8-1**. The standard deviation (s) is used to determine the appropriate number of significant digits. The result is rounded between 0.5s and 0.05s: 0.5s = 0.5000 x 0.3713 units = 0.1857 units; 0.05s = 0.05000 x 0.3713 units = 0.01857 units. The rounding unit is 0.1 units. A test result (e.g., 31.248 units) should therefore be reported as 31.2 units ***(14).***

Table 4.2.8-1 Confirming Significant Digits in Reported Test Results

Measurement Number	Result (in units)
1	31.233
2	31.766
3	30.899
4	31.444
5	31.002
6	30.776
Mean	31.1876
Standard Deviation (s)	0.3713
Rounding Unit	0.1

4.2.9 Validating Other Analytical Technologies

Some analytical procedures and the validation characteristics to be evaluated may not fit into the ICH Q2(R1) method type and any of the above listed categories. Examples include mass spectrometry, circular dichroism, NMR, measures of binding affinity, or procedures based on chemometric calculations applied to infra-red or fluorescent spectra, such as principal component analysis or partial least squares. In these cases, the reportable result (and specification) should be considered and used as a starting point to plan the AMV studies.

If the reportable result is quantitative, the principles outlined for validating quantitative test methods should be considered in the context of the intended purpose and applied as appropriate. If the reportable result is qualitative, the specificity of the method and its consistency should at minimum be evaluated.

If a qualitative method is to be used as a limit test, the validation design should verify the validity of that limit. If the reportable result is qualitative, but the underlying method has quantitative components, some of the validation characteristics associated with quantitative methods should also be considered.

4.3 Analytical Method Verification

A compendial procedure is considered validated when published in pharmacopoeias or other recognized sources. Revalidation is typically not expected, however, the suitability of the method with the intended product and/or material to be tested should be verified under actual conditions of use ***(26).*** The verification against preset acceptance criteria should provide acceptable results prior to using the compendial method to release product and/or material using representative samples and actual laboratory conditions. A similar approach may be acceptable for previously validated in-house methods, approved for a similar product, material, and/or manufacturing process. Changes in production process and/or formulation which may affect the analytical method performance may also be verified in a similar manner.

4.3.1 Verification Process

Verification should be conducted by the user to provide confidence that the approved method performs to expectations. The analytical performance characteristics should be evaluated based on preestablished acceptance criteria. These criteria depend on the intended use of the method. Upon completion of the verification studies, data should be compared to the preestablished acceptance criteria, and the conclusion about the suitability of the method under actual conditions of use with the intended product and/or material should be documented.

If the verification data do not pass the acceptance criteria, an investigation should be conducted similar to the process described in **Section 8.** The verification characteristics to be evaluated for each general method type are illustrated in **Table 4.3-1** for typical compendial method types.

Table 4.3-1 Verification Characteristics for Typical Compendial Method Types and Resulting Specifications

Method Types	Typical Specifications	Typical Minimum Verification Characteristics To be Evaluated
Identification	Yes/No Present/Absent Pass/Fail Consistent/Inconsistent	Selected or prepared relevant (blind) samples should be correctly identified to demonstrate specificity. Positive and, if applicable, negative identification should be demonstrated.
Impurity (Quantitative)	No More Than	Accuracy (against an acceptable reference standard) and repeatability and/or intermediate precision should be demonstrated using representative sample(s) below and above the QL.
Impurity (Limit)	Less Than	It should be demonstrated that impurity levels at or above the DL are reliable and can be detected in routine QC testing conditions.
Assay (Content, Potency, and/or Purity)	Range (for Content, Potency) No Less Than (for Purity)	Accuracy (against an acceptable reference standard) and repeatability and/or intermediate precision should be demonstrated using representative sample(s) within, below, and above the specifications.

4.3.2 Verification Requirements

When verifying that compendial and/or otherwise approved methods perform to expectations, the identified risks should dictate the design of the verification studies and the acceptance criteria. Similar risk concepts, as shown in **Section 3.1,** should be considered for the verification process and requirements.

Prospective verification may not be required for those compendial methods of low complexity, not used for product safety testing, while using acceptable calibration and control checks each time an assay is run. On the other hand, more extensive verification studies may be required to demonstrate the suitability of a microbiological safety test to be used for release and/or stability testing of a drug product administered by injection.

4.3.3 Retrospective Data

Retrospective data may be used to supplement the verification process for particular low-risk cases. Verification studies should ideally be prospective, using preset acceptance criteria. Similar to all validation studies, failing the verification criteria and fixing the problem prior to the routine implementation of the test method, should prevent inaccurate and/or unreliable data to be generated during GMP manufacturing.

4.4 AMV Documentation

AMV documents include a protocol which is approved prior to execution of the AMV studies, the raw data generated during the validation studies, and a report that summarizes the results. All of these documents should be readily available for review during and beyond the method life. This section describes key components of the method validation protocol and report.

4.4.1 AMV Protocol

The AMV protocol is an approved and controlled plan which contains all detailed executions steps, conditions, acceptance criteria, and justifications. Suggested protocol elements are shown in **Table 4.4.1-1.** Any relevant AMV-enabling historical data such as AMD/AMQ study reports and/or historical data generated

prior to AMV studies may be briefly summarized or referenced in the AMV protocol and/or report.

Table 4.4.1-1 Typical AMV Protocol Elements

Section No.	Section Title	Subsections
NA	*Protocol Approval*	Protocol Title; Signatures with Job Titles
NA	*List of Protocol Sections*	Table of Content; List of Figures (if applicable); List of Tables
1	*Introduction*	Intended Use and Sample(s) Description
2	*Method and Product/Process, Description*	Brief Description; (Target) Specifications
3	*Materials, Equipment, and Instrumentation*	Materials; Equipment; Instrumentation
4	*Historical Assay Performance*	Summary of Historical Data for Assay Control, samples, process capability, design space limits (if available), prior platform technology method performance (if applicable).
5	*Validation Characteristics, Design, and Acceptance Criteria*	Validation Prerequirements (if applicable); Validation Characteristics, Study Design, Sample Preparation, Acceptance Criteria
6	*Validation Execution Matrix*	Visualized Execution Process Map(s) and/or Execution Matrix Tables (see **Table 4.1.10-1)**
7	*Data Analysis*	Calculation Samples; Proposed Statistical Tests
8	*List of Procedures and Guidelines*	NA
9	*List of Attachments*	NA

4.4.2 AMV Report

The AMV report should be aligned with the protocol and should provide data and results for all required protocol studies. Suggested protocol elements are shown in **Table 4.4.2-1.**

Table 4.4.2-1 Typical AMV Report Elements

Section No.	Section Title	Subsections
NA	*Report Approval*	Report Title; Signatures with Job Titles
NA	*List of Report Sections*	Table of Content; List of Figures (if applicable); List of Tables
1	*Validation Summary*	NA
2	*Protocol Deviations*	NA
3	*Materials*	Materials; Product Lot Numbers; Reagent Lot Numbers
4	*Results and Discussion*	Validation Parameters; Results (Table and Text); Statistical Test Summaries; Discussion of Results
5	*Conclusions*	NA
6	*Data Analysis*	Calculation Samples (if not done in Protocol); Statistical Software; Sample Data Output(s)
7	*List of References*	NA
8	*List of Attachments*	NA
9	*AMV Matrix, Acceptance Criteria, and Results*	Table with Column Headings: Validation Characteristics, Validation Design, Sample Preparation, Acceptance Criteria, and Results

5.0 Analytical Method Transfer

Analytical Method Transfers (AMT) may occur at any point in the method and product life cycle. An AMT transfer may be associated with transfer of the entire manufacturing process during product development or after licensure, and/or be a portion of a larger technology transfer process. Or, an AMT may be required to implement the use of a new laboratory for quality control release and/or stability testing, either within or outside the company (e.g., contract lab). The AMT process should be similar for all post-validation cases.

The stages of an AMT include a preliminary evaluation and preparing the new laboratory to receive the test method, developing an approved method transfer protocol, and applying suitable statistical tools to analyze the results. The outcome is documented in a method transfer report.

5.1 Prerequisites to AMT

Prior to developing a specific method transfer protocol, the readiness of the receiving laboratory should be evaluated. Specifically, consideration should be given to the availability of required analytical and supporting equipment, software, critical reagents, standards, controls, and analysts who are skilled in the relevant analytical techniques as well as the qualification status of all materials, equipment, and analysts.

The test method procedure, method validation report, available historical data, and any prior method transfer reports should be reviewed to assess the readiness of the receiving laboratory prior to the actual transfer. If gaps are identified (for example, the receiving laboratory has a similar analytical instrument) a risk assessment should be performed, similar to **Sections 3.1 – 3.2,** before execution of the formal transfer studies.

Some methods will also require attention to the lab environment (humidity and temperature) as part of the prerequisite review. Shipment and receiving procedures are needed to allow transfer of critical reagents, standards, and samples between laboratories. Incorporation of the test method procedure into the receiving laboratories quality system is also part of the transfer process.

Arrangements should be made for the originating laboratory to provide hands-on training in the specific test method to analysts at the receiving laboratory, if needed. Alternatively, hands-on training could be provided at the originating laboratory. The type and amount of training needed will vary depending on the analytical method transferred and the existing experience of the receiving lab and its personnel. Successful completion of training should be documented as per the quality system requirement prior to executing the method transfer protocol. In addition, evaluation of the capability of the receiving laboratory to execute the system suitability requirements of the method successfully during the training is recommended.

If the receiving laboratory is at a contract site, an audit should be performed and a quality agreement should be in place prior to analytical method transfer(s) execution. A review of the contract lab AMT process procedure(s) should be part of the quality audit.

For all AMTs, the responsibilities between sending and receiving laboratories should be established. **Table 5.1-1** lists the suggested responsibilities for each laboratory. The quality and/or service agreement(s) should clarify all conditions and responsibilities as they may vary depending on the situation. In addition to the preparation and sharing of samples, critical reagents and standards to be used during the AMT studies, some continuous post-AMT testing (monitoring) should also be considered. **Table 5.1-1** provides some examples how tasks and responsibilities could be shared by both laboratories during the AMT process.

Table 5.1-1 Suggested AMT Responsibility Matrix

Lab	Suggested Responsibilities
Sending lab	– Feasibility/readiness assessment – Compile QC/process data – Organize training if required – Establish the transfer package – Write transfer protocol based on requirements of both labs and knowledge of method prior to transfer – Establish protocol acceptance criteria – Allocate resources for training and transfer study – Provide critical reagents and samples – Provide troubleshooting support – Approve the transfer report
Receiving lab	– Review the transfer package – Define the transfer process including training requirements – Inform the donor lab on potential issues identified (such as different suppliers in critical equipments) – Allocate resources for training and transfer study – Analyze transfer data – Write the transfer report – Inform the donor lab of the outcome of the transfer – Approve the transfer report

5.2 General AMT Strategy

The strategy used for an individual method to be transferred and/or to support a product transfer can vary. Several possible options, as illustrated in USP <1224> Transfer of Analytical Procedures ***(27),*** are shown below and others may also be acceptable in certain situations.

Co-validation – Sending and receiving laboratories participate in the AMV study execution. This may be used early in the life cycle of a test method when appropriate.

Comparative study – AMT study performed concurrently by sending and receiving laboratories. Acceptance criteria determine the equivalence of the two laboratories. The sending laboratory typically has collected a significant amount of historical data for test method performance results in addition to test results for the samples to be tested at the receiving laboratory. Historical and validation data from the sending laboratory may be used when appropriate for parts of the method transfer study and may not have to be repeated. Acceptance criteria for the AMT should be derived following the process and conditions as illustrated in **Section 3.2.** Acceptance criteria could be set based on previous validation/qualification studies and/or recent routine QC testing data with respect to the relevant product or material specifications.

Performance Verification – The receiving laboratory may already perform the method for a similar product or for another type of sample for the same product. In this case, a formal method transfer

may not be required. Any prospective study could be designed similar to a compendial method verification study (**Table 4.3-1**).

Because the intent of this section is to describe a typical post-validation transfer, a comparative study model is further described below.

5.3 Design of Comparative (AMT) Test Studies

The AMT protocol should include a study design specifying method parameters to compare, samples to test, justified acceptance criteria, and the statistical methodology to evaluate the results.

5.3.1 Selecting AMT Performance Characteristics

The actual intended purpose of the method should be used to justify the rationale of the study design and acceptance criteria for each method transfer. **Table 5.3.1-1** is an example of performance characteristics to be compared between laboratories for different types of methods. Other performance characteristics covered during the validation studies may also be considered.

Table 5.3.1-1 Examples of Method Types and AMT Performance Characteristics

Type of Methods	AMT Performance Characteristics Examples
Identity tests	System suitability, specificity, qualitative comparison (if applicable)
Impurities (quantitative) – process- and/or product-related	System suitability, precision and accuracy; consider several concentration levels: minimum reportable quantity and / or Quantitation Limit(s) and 120% of the product specification; stability samples may need to be included to assess stability-indicating capabilities, as relevant
Impurities (qualitative, limit)	System suitability, Detection Limit(s)
Assay – content and potency	System suitability, precision and accuracy, range, and stability samples may need to be included to assess stability-indicating capabilities, as relevant

5.3.2 Sample Selection and AMT Study Design

Representative critical in-process sample types as well as drug substance (DS) and drug product (DP) should be selected as appropriate for the application and purpose of the method. When comparing stability-indicating methods, degraded samples could be directly compared by both laboratories. Whenever sample preparation specific to the AMT studies are done, such as different spiking levels of samples, the preparation itself may significantly contribute to variation in method transfer results. It is therefore important to control this process to reduce this unexpected variation and/or bias during the method transfer process.

It may also be beneficial to use multiple batches of samples and/or material representing typical concentration levels (or bracketing for products of different strengths) and/or matrix variation to ensure that the analytical method performance remains sufficient over these ranges. These extended AMT results may provide additional information whether both labs can produce similar accuracy (matching) and precision (reliability) results over the potential range of results expected.

A sufficient number of samples and testing runs should be executed to establish equivalence between the two laboratories. The ability to detect a difference or establish confidence that no difference exists is directly dependent on the number of determinations (number of results from independent runs) for each laboratory. Two approaches for choosing the sample sizes could be envisaged based on the method complexity and its known variability.

The AMT study design can consist of a "fixed" execution matrix, similar to that illustrated in the ISPE Technology Transfer Guide (**28**), and/or "variable" execution matrix. A fixed AMT execution matrix does not integrate known test method result variation and has therefore an identical set of comparative data generated between both laboratories for each method transfer executed. A fixed Matrix can be more advantageous when transferring multiple products to/from multiple locations. The fixed number of replicates and acceptance criteria are set for the relative difference between means found at both sides or by Equivalence testing using two one-sided t-Test (TOST, see **Table 5.3.2-1** below). The study typically addresses at least two independent factors (e.g. "analysts" and/or "days") known from the AMV studies to (potentially) impact routine test result reliability. Intermediate Precision at both laboratories can be evaluated from this data set, however, when a more detailed result interpretation is desired at the receiving laboratory, a more extensive set-up may be more appropriate.

A variable execution matrix does consider test method result variation and may require a larger data comparison set for highly variable test methods. The selection of the AMT study design should be considered for a given situation. For example, a variable execution matrix may be advantageous when transferring bioassays with a relatively high degree of test result variation. On the other hand, when simultaneously transferring multiple assays in support of a product manufacturing transfer, a fixed execution matrix may be the most effective process.

Table 5.3.2-1 Examples of AMT Execution Matrices and Acceptance Criteria

Method Type	AMT Execution Matrix Examples	AMT Protocol Acceptance Criteria Examples
Identity	Results for multiple positive and negative samples should be compared when comparing Specificity (differentiation capability). Blind sample testing may be required for non-automated identification systems.	System suitability met, similar or identical differentiation capability should be demonstrated.
Impurities (quantitative) – process- and/or product-related	Two operators and/or instruments on different days, three batches in duplicate; consider spiking at different levels for confirming precision, accuracy and Quantitation Limit(s).	System suitability met, Quantitation Limit(s) confirmed, TOST difference of less than or equal to 10% with 95% confidence for moderately high levels of impurities, or, absolute difference of the means between laboratories between ± 25% for low levels of impurities. Note: Results for different batches may not be pooled unless normalization prior to comparing can be justified.
Impurities (qualitative, limit)	Results for multiple samples below and above the Detection Limit(s) should be compared.	System suitability met, similar Detection Limit's should be demonstrated.
Content uniformity, purity, and/or potency	Two operators and/or instruments on different days, at least one batch, number of occasions according to uniformity demonstration criteria at each laboratory; consider bracketing with batches of multiple strengths.	System suitability met, TOST difference of less than or equal to 3% with 95% confidence, or, absolute difference of the means between laboratories between ± 3% with comparable data variance. **Note:** Results for different batches may not be pooled unless normalization prior to comparing can be justified.

5.3.2.1 Specific AMT Study Design for Highly Variable Methods

For highly variable methods, such as certain bioassays, the use of fixed matrices may not be adequate. An appropriate sample size can be determined using the risk-based approach outlined in USP's proposed general chapter PF *35*(2) <1033> *Biological Assay Validation* which takes into account both Type I and Type II errors ***(29).*** In hypothesis testing, the Type I error represents the risk of rejecting a parameter that is actually satisfactory while the Type II error represents the risk of accepting a parameter that is actually unsatisfactory (as shown in **Table 5.3.2.1-1).**

Table 5.3.2.1-1 Type I and Type II errors

	H_0 is true	H_0 is false
H_0 is accepted	√	Type II error
H_0 is rejected	Type I error	√

The probability of committing a Type I error is α (the specified significance level) and the probability of committing a Type II error is β (generally not specified nor known). Both types of error can be reduced simultaneously by increasing *n*.

USP PF *35*(2) <1033> formula for calculating the sample size (*n*) needed for demonstrating equivalence to acceptance criteria Θ (accepted difference) takes into account both type I (α) and type II (β) error using the estimate of standard deviation for Intermediate Precision (s_{IP}) and t distribution points, or substituting z (the percentiles of the standard normal distribution) for t.

[Equation 2]

$$n \geq \frac{2\times(t_{\alpha/2,\,n-1} + t_{\beta,\,n-1})2\text{x}s^2_{IP}}{\Theta^2} \approx \frac{2\times(z_{\alpha/2} + z_{\beta})2\times s^2_{IP}}{\Theta^2}$$

Based on the sample size *n*, the study plan should be appropriately designed so that at least two independent factors (e.g., analysts and/or days) known to (potentially) impact test method results, are investigated during the transfer. Statistical Equivalence testing is usually applied to confirm transfer acceptance to preset criteria.

Table 5.3.2.1-2 General AMT Design Parameters and Considerations

AMT Design Parameter	Suggested Considerations
How many representative batches – Matrix (number of different sample types and/or batches to be evaluated)	Two or three batches bracketing expected active protein concentration range could be used. The selected materials should be representative of routine samples. Retain samples, reference standards, samples at the extremes of acceptance limits, stability samples, and/or spiked samples should be used depending on the situation. For impurity tests, samples may be spiked or degraded so that the level of the impurity is below and/or above the QL (and/or specification limit). If samples with a measurable impurity level are not available, then it may be necessary to prepare spiked samples to evaluate the accuracy and precision of measurable amounts of impurity/degradation levels during the AMT studies. If there are differences in the formulation, adequate testing of the range of formulation differences should be included. The rationale for the selection of representative AMT samples should be documented in the AMT protocol.
How many replicates per sample and lab? (number of independent runs)	The number of replicates depends on the Repeatability and Intermediate Precision performance of the method to be transferred and the desired confidence level(s) for meeting product specifications. The AMV report and other related data sources (for example, routine test results) should be reviewed.
How many Intermediate Precision variability factors ?	At least two critical factors should be selected based on prior knowledge of which factor(s) may have the greatest expected impact on test result variation.

5.4 Acceptance Criteria and Statistical Evaluation

5.4.1 Acceptance Criteria for AMT Study

Acceptance criteria should be established and justified for the allowed difference(s) between the originating and receiving laboratories prior to the transfer. Risk assessments following similar concepts as those illustrated in **Section 3.1** should be performed when establishing acceptance criteria. The intended statistical evaluation methodology should be considered. Acceptable differences between laboratories for the method performance characteristics of quantitative methods such as Accuracy and Intermediate Precision should be estimated based on historical data and/or previous AMV protocols/reports with respect to the specifications.

Acceptance criteria examples, stated in the **Table 5.3.2-1,** are based on typical analytical procedures used to test pharmaceuticals. The use of wider acceptance criteria may be justified for the testing of biotechnological products whenever specifications are relatively wide. As a general rule, acceptance criteria (accepted mean difference between two laboratories) can be equal to or greater than the Intermediate Precision (s_{IP}) and should be less than the lower/upper specification limits (LSL, USL).

[Equation 3]

$$\left[-\frac{s_{IP}}{\sqrt{n}}, +\frac{s_{IP}}{\sqrt{n}}\right] \leq [-\Theta, +\Theta] \leq [LSL, USL]$$

Alternatively, the validation results and/or routine testing results could be used to set acceptance criteria ***(30).*** When statistically deriving the acceptance criteria from Intermediate Precision, the upper limit of the one-sided 80% confidence interval of the Intermediate Precision found during the validation is estimated to take the Reproducibility (inter-laboratory) factor into consideration and is used in the determination of the acceptance criteria Θ.

[Equation 4]

$$\Theta = s^{*} \times (t_{\alpha,2n-2} + t_{\beta,2n-2}) \times \sqrt{\frac{2}{n}}, \text{where } s^{*} = s \times \sqrt{\frac{n-1}{\chi^{2}_{(\gamma,n-1)}}}$$

Other options to set risk-based acceptance criteria may also be acceptable as long as the concepts are similar and justified as those used in **Section 3.2.** For the AMT results of any method to be acceptable, all quality control criteria, defined in the analytical procedure, should be met. For example, the system suitability requirements should be satisfied, the instrumentation performance should be within the acceptable range(s), and all test results should be within the established acceptance criteria.

The demonstration of equivalence in mean results (accuracy and/or matching) and a similar precision (Intermediate Precision) performance between the laboratories is of primary interest in evaluation of quantitative methods. Additional validation characteristics comparisons (such as QL or DL) may be considered for particular method types and their intended use. Statistical tests can be used to demonstrate equivalence between laboratories.

5.4.2 Statistical Tests for AMT Studies

Inferential statistics and Hypothesis Testing are commonly used to support method transfer acceptance or rejection. Traditional Hypothesis Testing such as mean comparison by T-test or variance comparison by F-test can generate results indicating either a statistically significant or statistically non-significant difference exists for the transfer data set regardless of practical significance. Equivalence testing by TOST is generally applicable in most cases ***(31).*** The Hypothesis Testing and acceptance criteria should be selected and justified to minimize the risk(s) of failing the formal study due to insignificant differences.

Equivalence testing, by TOST is statistically satisfactory if the confidence interval for the difference in means between the two laboratories falls within an acceptable interval [–Θ, + Θ] ***(32).*** The interval should define the largest difference that could be accepted between the laboratories while not significantly impacting the test results. Specifications, internal control limits associated with the analytical method, and overall method precision should be considered when determining the acceptable interval, as explained in the above section. When comparing results from two laboratories, the interval is centered onto zero, reflecting the fact that there is no bias between both laboratories. Based on the two sets of results and the pooled standard deviation (s_p) from the two laboratories, a confidence interval is calculated for the difference in sample means. The null hypothesis that the means are not equivalent is rejected once the confidence interval is strictly found within the acceptance interval. The two sets of results are therefore considered as equivalent.

[Equation 5]

$$-\Theta < \left[(x_1 - x_2) \pm t_{1-2\,\alpha,n1+n2} - 2 \times S_p \times \sqrt{\frac{1}{n1} + \frac{1}{n2}} \right] < +\Theta$$

The pooled standard deviation (s_p) is calculated using the following equation:

[Equation 6]

$$s_p = \sqrt{\frac{(\mathrm{n1} - 1) \times s_1^2 + (\mathrm{n2} - 1) \times s_2^2}{n_1 + n_2 - 2}}$$

Boxplots, also known as box-and-whisker plots, can be used as visual aids to compare results. Boxplot examples are shown in **Figures 5.8.1** and **5.8.2.** Boxplots are convenient non-parametric visual tools for comparing two sets of results through descriptive parameters, including the minimum and maximum values, the lower (25^{th} percentile) and upper (75^{th} percentile) quartiles and the median. The plots may be helpful in indicating the degree of dispersion and skewness of results, and may help in identifying potential outlying results. Of great importance is to at least verify visually that data distribute normally around their mean, since it is often a prerequisite in hypothesis testing (including TOST).

5.5 Sample Preparation

Test samples used in AMT studies should be carefully prepared, shipped, and stored to avoid differences in test results at the time of testing in both laboratories. Besides the sample preparation, shipping, and storage conditions, sample homogeneity and sample stability should also be considered for the AMT studies. Some additional considerations are listed below.

- The set(s) of sample preparations should be as uniform as possible.
- Suitable reference and/or control material should be selected and included in each single analytical run.
- Reference material should be sufficiently characterized.
- Sufficient sample and reference material aliquots should be prepared to allow for additional testing in case invalid test results are generated.
- All samples and reagents should be accompanied by a COA or other suitable documentation.
- Distribution and storage conditions should be defined based on expected stability of all material to be tested.

5.6 Deviations and Failures

Any deviation to the approved AMT protocol should be properly justified and approved. If any of the acceptance criteria stated in the protocol are not met during the execution of the AMT study, an investigation should be performed and proper corrective and preventative actions implemented. More detail on the failure investigation process is provided in **Section 8.0.**

5.6.1 Invalid Assays

Assays which do not meet system suitability criteria specified in the test method are not included in the analysis of results for comparison to the protocol acceptance criteria. Additional assays should be performed to replace the invalid ones. However, the invalidated runs must be stated in the AMT report with the rationale for their exclusion.

5.6.2 Handling of Outlaying Results and Retesting

If a test result is a suspected outlier, statistical outlier testing, using a desired confidence level can be performed. Outlier tests which can be used are the Generalized Extreme Studentized Deviate (ESD) test, Dixon-Type test, or Hampel's Rule ***(33, 34).*** Alternative outlier tests may also be acceptable depending on the situation.

The confirmation of an outlier result is by itself not sufficient to cause an analytical result to be considered invalid. The confirmation should assist in the root cause investigation but should not be used in lieu of any investigations. Investigation questions similar to those as raised in **Section 8.0** should be answered as part of this investigation.

5.6.3 AMT Study Extension

In case the initial sample size results (N_1) have generated confidence interval(s) too wide for a clear pass/fail conclusion, the study could be extended whenever no other apparent cause exists with an additional set(s) (N_2) of independent runs. All results should then be pooled ($N_1 + N_2$) before final interpretation. If this option is not already considered in the original AMT protocol, an AMT protocol amendment should be approved before execution of additional data sets.

5.7 AMT Documentation

Similar to AMV documents, AMT processes are documented through AMT protocols and AMT reports. The AMT protocol typically consists of the following sections as illustrated in **Table 5.7-1** below.

Table 5.7-1 Typical AMT Protocol Sections

Section No.	Section Title	Subsections
NA	*Protocol Approval*	Protocol Title; Signatures with Job Titles (and Responsibilities)
NA	*List of Protocol Sections*	Table of Content; List of Figures (if applicable); List of Tables
1	*Introduction*	Intended Use and Sample(s) Description
2	*Method and Product/Process, Description*	Brief Description; (Target) Specifications
3	*Samples, Materials, Equipment, and Instrumentation*	Sample Preparation and Storage; Materials; Equipment; Instrumentation, Personnel
4	*Historical Assay Performance*	Summary of Historical Data for Assay Control, samples, process capability, design space limits (if available), prior Analytical Platform Technology method performance (if applicable).
5	*AMT Characteristics and Design, and Acceptance Criteria*	AMT Characteristics, Statistics, Acceptance Criteria, Justification(s)
6	*AMT Execution Matrix*	Visualized Execution Process Map(s) and/or Execution Matrix Tables
7	*Data Analysis*	Calculation Samples; Proposed Statistical Tests
8	*List of Procedures, References, and Guidelines*	SOP(s), AMV Protocol/Report(s); Other References
9	*List of Attachments*	NA

The AMT report describes the results of performance of the protocol and compares these results to the acceptance criteria and draws a conclusion on acceptability of the transfer.

5.8 AMT Example

A validated analytical method for potency is to be transferred from the original QC laboratory to another QC laboratory to release drug product (DP). The analytical method generates potency (dose)

results for lyophilized DP. The vials are available in three nominal doses between 500 – 2000 IU/vial using an identical formulation. Release testing is performed using three replicate preparations from each of three vials. Before analysis the content of a vial is reconstituted with 5.0 mL of WFI water and the potency is measured in IU/mL (100 – 400 IU/mL). The samples and a product-specific reference standard are prepared similarly. The analytical method procedure and statistical evaluation are performed with the parallel-line concept. **Table 5.8-1** lists the AMT study design details.

Table 5.8-1 AMT Study Design

Characteristics Evaluated	Accuracy/Matching: The relative difference between lab means should at 90% confidence not be less than $-\Theta = 10\%$ and not more than $+\Theta = 10\%$. The 10% difference limit was set with consideration of product specification. Intermediate Precision: RSD 6 % for all sample types, with appropriate homoscedasticity throughout the potency range (from validation) This means that any RSD from a sample of n=8 should not exceed 9.43% (1)
Number of Replicates	$N_{replicates}$ = at least 23 independent replicates(2) The confidence interval for the lab-to-lab difference for N determinations to less than the [10%, +10%]. As above the 10% difference limit was set with consideration of product specification.
Samples to test	$N_{level} = 3$ The range of potency/dosing results is covered by: Lowest dose 500 IU/vial or 100 IU/mL Medium dose 1000 IU/vial or 200 IU/mL Highest dose 2000 IU/vial or 400 IU/mL
Testing design, each sample	Number of operators, n = 2 Number of days, n = 2 Number of replicates per day per operator, n = 2 N = 8 in each lab for each of n= 3 potency levels Total N_{Total} = 24 individual observations will be recorded for each laboratory. N=24 individual observations are needed as N=23 is the minimum number of replicates calculated.

1. Upper 80 % confidence limit for standard deviations. Excel function: = 6* SQRT((8-1)/(CHIINV(1-0.1,7)))
2. Using equation from footnote (2) in **Table 5.3.2-1,** $\alpha = \beta = 0.05$, z_α from EXCEL function NORMSINV(α), Θ expressed in % =10 and s_{IP} in RSD = 9.43

[Equation 7]

$$n \geq \frac{2 \times (z_{\alpha/2} + z_\beta)2 \times s^2_{IP}}{\Theta^2} = \frac{2 \times (-1.96 - 1.6449)^2 \times 9.43^2}{10^2} = 23$$

When bracketing potency with reference samples of different potencies, it is often helpful to express results in recovery percentage so that results are normalized and can be pooled over the potency range. The recovery percentage is expressed as the ratio of the response measured in the reference sample from the calibration curve over the nominal theoretical titer (see **Table 5.8-2).** The limits of 90% confidence interval on the laboratory difference are then compared to the acceptance criteria [–Θ, +Θ] to determine the pass/fail status of the AMT according to equations described in the **Section 5.3.3-2.**

Table 5.8-2 AMT Transfer Results

Theoretical Potency Level in IU/mL	Operator	Day	Replicate	Sending lab		Receiving lab	
				Experimental Potency in IU/mL	%Recovery vs. Theoretical Potency	Experimental Potency in IU/mL	%Recovery vs. Theoretical Potency
100	1	1	1	103	103.0	95	95.0
100	1	1	2	104	104.0	99	99.0
100	1	2	1	108	108.0	104	104.0
100	1	2	2	101	101.0	103	103.0
100	2	1	1	94	94.0	93	93.0
100	2	1	2	99	99.0	96	96.0
100	2	2	1	102	102.0	92	92.0
100	2	2	2	104	104.0	100	100.0
200	1	1	1	212	106.0	208	104.0
200	1	1	2	208	104.0	192	96.0
200	1	2	1	191	95.5	199	99.5
200	1	2	2	201	100.5	195	97.5
200	2	1	1	204	102.0	208	104.0
200	2	1	2	206	103.0	211	105.5
200	2	2	1	198	99.0	203	101.5
200	2	2	2	200	100.0	183	91.5
400	1	1	1	375	93.8	383	95.8
400	1	1	2	401	100.3	401	100.3
400	1	2	1	408	102.0	389	97.3
400	1	2	2	388	97.0	391	97.8
400	2	1	1	402	100.5	408	102.0
400	2	1	2	415	103.8	421	105.3
400	2	2	1	406	101.5	415	103.8
400	2	2	2	410	102.5	403	100.8
TOST with acceptance criteria [–10%, +10%][1]				N1	24	N2	24
				Mean1	101.1	Mean2	99.3
				SD	3.5	SD	4.2
				RSD	3.4	RSD	4.2
				Pooled SD[2]	3.9		
				Mean1-Mean2	1.8		
				t-value	1.679		
				Upper 90% CI limit	**4 (3.6)**		
				Lower 90% CI limit	**0 (–0.1)**		
Transfer Acceptance Conclusion					**Pass**		

1. Raw data was used unrounded. Upper and lower 90% CIs were calculated using equation in **Section 5.4.2.**
2. Equal variance was confirmed using an F-test to justify the pooling of the standard deviation.

Visual examination of the AMT results using a boxplot, as illustrated in **Figure 5.8-1,** is helpful to assess the test result variation within each laboratory in relationship to the mean difference(s) The boxes represent the 25th – 75th percentile distribution of the results between the two laboratories. Medians (line in the box) and means (cross in the box) are approximately centered while the medians are equidistant from the box hinges, providing a visual indication for a normal data distribution(s) among data points within each laboratory set. One potential outlier (lower open circle outside of the whiskers) is observed in the sending lab, however, this does not change the overall interpretation for the demonstration of lab-to-lab equivalence. The variation in the test results (wider 25th – 75th percentile boxes) appears to be higher in the receiving laboratory. This can often be attributed to less test method execution experience at the receiving lab at the time of transfer.

Figure 5.8-1 Graphical Representation of Potency Results Per Potency Level Between Laboratories

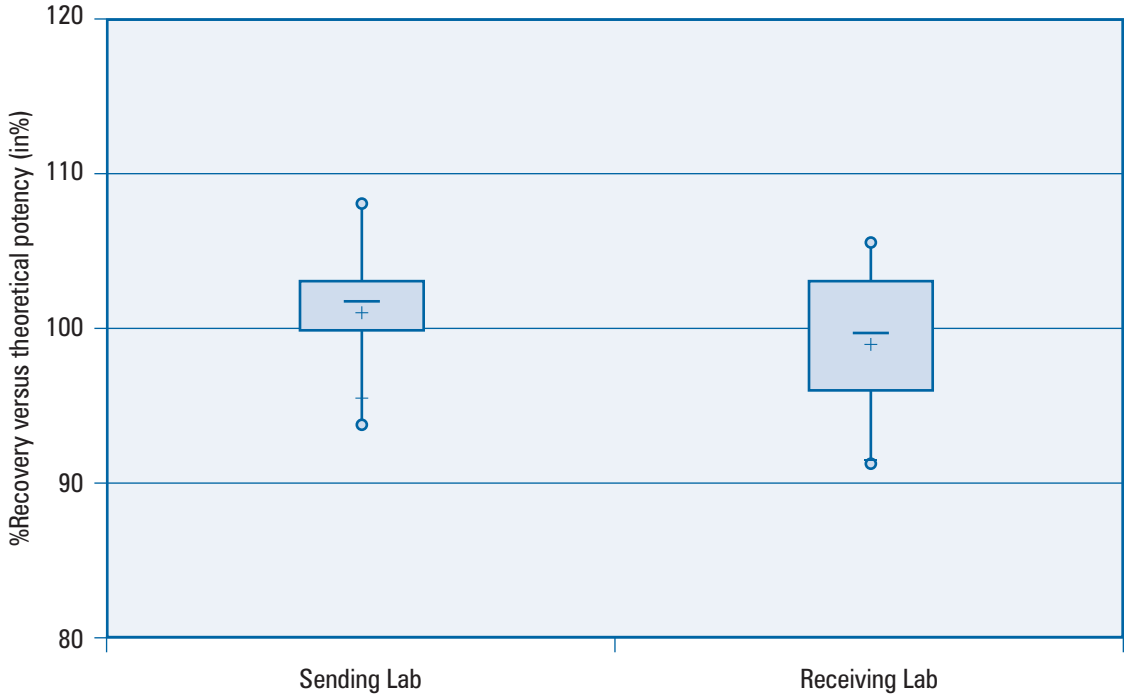
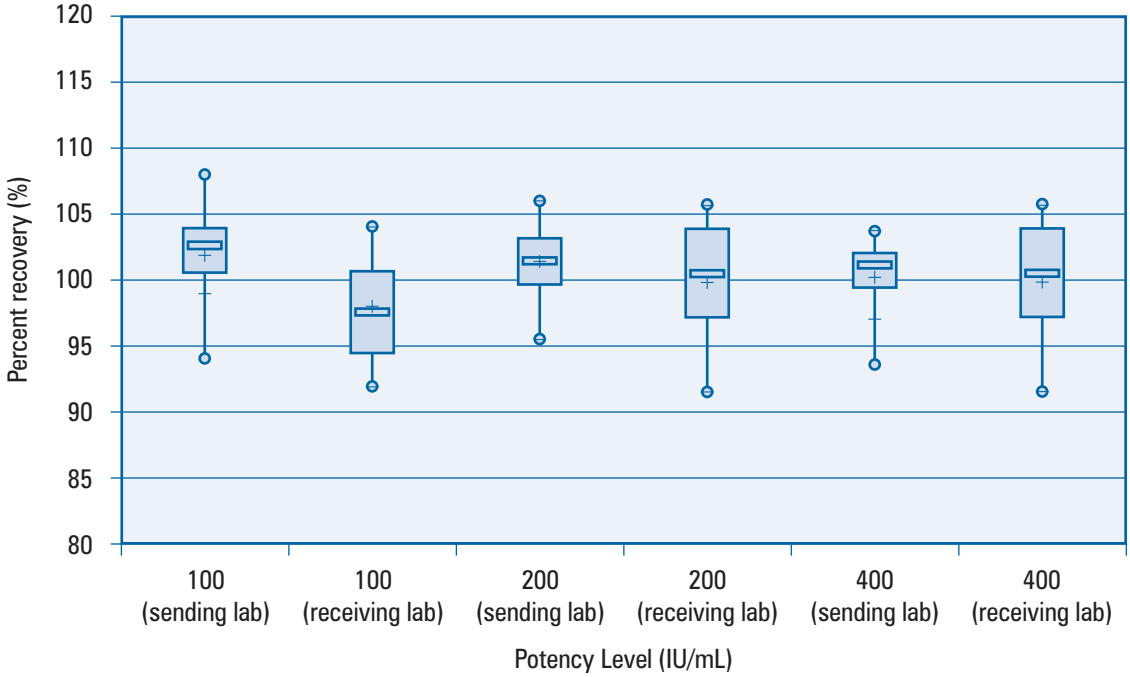


Figure 5.8-2 illustrates the AMT results for each of the three potency levels tested between both laboratories. Plotting the three potency levels within one figure allows for a visual assessment of variation homoscedasticity within each laboratory over the potency level range (100 – 400 IU/mL). The boxes represent the mean differences at each level, and the error bars represent the corresponding 90% confidence interval(s) for the mean difference(s). This boxplot also provides a visual assessment of acceptable accuracy/matching between laboratories for each potency level tested.

Figure 5.8-2 Graphical Representation of the Combined Percent Recoveries Between Laboratories for All Three Concentration Levels



Using both boxplots (**Figures 5.8-1** and **5.8-2**), the equivalence of the two laboratories could be demonstrated.

5.9 AMT Continuum

The AMT process should include continuous technical support from the sending laboratory. In particular, critical and/or specific reagents and material should be made available to the receiving laboratory before and after the AMT studies, if needed. Technical support from the sending laboratory should be provided to the receiving laboratory in the event of deviations, excessive invalid assay runs, and/or unexpected results.

Some post-AMT trending data should be assessed at the receiving laboratory to ensure a continuous validation state. Ensuring a post-transfer data continuum is conceptually similar to continuous validation activities and is discussed in **Section 7.**

6.0 Analytical Method Comparability

Formal AMC studies apply when an approved test method is to be changed for a new test method. The AMC data are an important part of the AMV results for the new method, as regulatory approval for the use of the changed test method is contingent upon the submitted AMC studies.

6.1 Replacing Analytical Methods

An equal or improved analytical method performance for the new or candidate test method versus the approved one should be demonstrated by AMC studies. The AMC studies could be included as part of the formal AMV protocol or could simply be executed under a separate protocol after the AMV was completed. Performing a separate method comparability study after AMV completion has the advantage that, if the AMV results reveal that a method was not optimized, time and effort will be saved by holding off the comparability studies until the new method is ready for AMC studies.

Table 6.1-1 lists the suggested performance characteristics to be statistically compared for the AMC studies each test method category. The four general ICH Q2(R1) test method categories can be grouped into two greater categories: qualitative and quantitative test methods. A qualitative test method provides qualitative results (pass/fail, yes/no, or results reported simply as "less than" some action or specification level), whereas a quantitative test method provides results reported in the same units as the specifications.

Qualitative test methods need not be accurate or precise, but they should be specific for the analyte tested. It is therefore critical for qualitative methods to provide high percentages of positive results for positive samples and vice versa, high percentages of negative results for negative samples. For qualitative limit tests, a low DL is desirable as it increases the likelihood for observing positive results at low analyte concentrations.

For all quantitative methods, the method performance characteristics accuracy and precision (intermediate precision) should be compared. Assuming that both methods were properly validated individually, it remains a regulatory and operational concern whether results will be expected to change overall by shifting (change in "accuracy") or by a potential increase in day-to-day variance ("intermediate precision").

A risk assessment should be performed to determine whether additional AMV characteristics, besides those suggested in **Table 6.1-1,** may need to be evaluated and/or compared. Assuming that the new test method is properly validated and the relevant risk(s) is/are low, a comparison of additional method performance characteristics may not be required.

Table 6.1-1 Suggested Statistics to Assess AMC for Each Method Performance Characteristic ***(14)***

ICH Q2(R1) Category	Identification Test (Qualitative)	Limit Test (Qualitative)	Limit Test (Quantitative)	Potency or Content (Purity or Range) (Quantitative)
Accuracy	Not Required	Not Required	T-test	T-test
Intermediate Precision	Not Required	Not Required	ANOVA, mixed linear model, or other F-test statistics	ANOVA, mixed linear model, or other F-test statistics
Specificity	Probability and/or Chi-Squared for Number of Correct Observations	Probability and/or Chi-Squared for Number of Correct Observations	Not Required	Not Required
Detection Limit	Not Required	Depends on how DL was established. Probability calculations may be used	Not Required	Not Required

AMC studies executed during or after the AMV studies for the new test method should be formally completed under a protocol using preset acceptance criteria. To obtain regulatory approval for changed test methods based on the AMC data, "comparable" means that new test methods should perform equal to or better than the to-be-replaced methods (compendial, officially recognized, or approved methods) with respect to the relevant method performance characteristics.

6.2 Demonstrating AMC in Post-Validation Studies

Because of the conceptual similarity between pre- and post-validation AMC studies, only the formal post-validation demonstration of performance comparability is discussed. The type of comparison used in the protocol is dependent on the stated objective of the comparability protocol. In a few cases the objective may be to demonstrate the superiority of the new method relative to the current method, however, in most cases it may be sufficient to demonstrate noninferiority or equivalence.

The categories of noninferiority, equivalence, and superiority, are described in ICH E9 and can be used for the comparison of method performance ***(10,35).*** The suggested statistical tools (**Table 6.1-1**) used for qualitative and quantitative comparisons are described with their possible outcomes. Replacing a qualitative test with a quantitative test, and vice versa, may require an adjustment of existing specifications, and should be treated on a case-by-case scenario.

The prespecified limits to establish equivalence and/or noninferiority should be justified prior to the execution of the formal AMC studies. The noninferiority limit (–Delta) is a one-sided limit whereas equivalence limit is a two-sided limit (–Delta and +Delta) as illustrated in **Fig. 6.3.1-1**. The superiority limit is typically set at the no-difference point and does therefore not require additional justification.

6.2.1 Qualitative Tests

All qualitative tests should contain, at minimum, a comparison of positive-to-fail ratios of spiked (low) analyte concentrations. This will ensure a comparable level of specificity of both methods. For the DL, both hit-to-miss ratios can be compared at very low analyte concentrations using probability statistics.

When comparing qualitative data, noninferiority or superiority models should be used and three possible outcomes are illustrated below.

- Inferiority. A particular performance characteristic compared provides significantly inferior results for the current method, therefore failing to demonstrate AMC.
- Noninferiority. The new method performs at a comparable level. The new method could be superior, equivalent, or insignificantly inferior. All three outcomes are acceptable outcomes to demonstrate noninferiority.
- Superiority. The new test method is superior. When testing for superiority, only this outcome is acceptable.

6.2.2 Quantitative Tests

For all quantitative methods, the method performance characteristics accuracy and precision (intermediate precision) should be compared. Similar to the comparison of qualitative data, the comparison of intermediate precision for quantitative tests could have three acceptable outcomes (noninferiority, equivalence, or superiority). Depending on the prespecified allowable difference, a significant shift in results may require a change in the release specifications or other possible adjustments before the new method can be used for release testing. The demonstration of comparable accuracy or "result matching" of a method will therefore require the use of an equivalence model.

When comparing quantitative data for accuracy, two possible outcomes are illustrated below:

- No equivalence for accuracy. The observed statistical difference (e.g., 90% confidence intervals) is not within the predefined acceptance criteria. The new method may be acceptable if specifications changes are justifiable or other adjustments can be made.
- Equivalence. The statistical difference between both methods (e.g., 90% confidence intervals) is completely enclosed in the acceptance criteria, i.e. the new method performs at a comparable level.

6.3 Design of AMC Study

The method comparability protocol is to establish or demonstrate that the new method intended to replace the current method is indeed an acceptable substitute. The regulatory implications of the method change necessitate a careful evaluation of the impact of the method change on specifications—not only release specifications, but also end of shelf-life specifications in the case of changes to stability-indicating methods.

The type of comparison utilized could be different for each of the method performance characteristics studied. For example an equivalence test may be utilized for accuracy while a noninferiority test may be used for intermediate precision. The choice of samples used in the comparability studies is dependent on the purpose for use of the method and the objective of the protocol. For stability-indicating methods, stressed samples, and retains from stability pulls may be included in the comparability study. When selecting samples for comparison, the product specification and process history should be considered.

When defining the AMC protocol margins for the maximum allowable differences. The number of samples to be compared for each method performance characteristic may depend on the type of comparison category and the allowed difference. The potential impact on the specifications should be taken into consideration when establishing the equivalence margins (+Delta, –Delta).

To statistically demonstrate comparable accuracy, two-sided confidence intervals (CI) can be used for equivalence testing and equivalence is inferred if the 90% CI falls within the prespecified equivalence margins.

6.3.1 Application and Acceptance Criteria

The proper application of the three AMC categories should be understood with respect to each method performance characteristic. In some cases, AMC data sets can be used for the evaluation of multiple method performance characteristics. For example, when accuracy is compared using results generated side-by-side by both methods under "intermediate precision" conditions to demonstrate equivalence of the methods; each method's intermediate precision can also be simultaneously compared.

Illustrated in **Figure 6.3.2.1-1,** to demonstrate noninferiority, the 95% confidence interval for the mean difference of the new test method versus existing test method must fall to the right of a prespecified noninferiority margin. Similarly for superiority, the 95% confidence interval must fall entirely to the right of 0 (no difference). For equivalence, the 90% confidence interval must fall within the prespecified lower and upper equivalence margins (+Delta, –Delta).

When using one of the suggested three AMC categories, two important points should be considered.

1. The chosen comparison category should be explained and justified. For example, a noninferiority test may be appropriate to demonstrate that the candidate method is non-inferior, if all outcomes (noninferiority, equivalence, and superiority) are acceptable, and if the new method is superior in other aspects such as real-time testing/results, increased sampling, or increased reliability. An AMC protocol should include a detailed design of experiments to be performed and all statistical test(s) to be used.

2. The prespecified maximum allowable difference(s) should be derived and justified. The difference limit(s) should strike a balance among possible opposing incentives. Delta should be derived similar to the acceptance criteria for AMV protocols.

6.3.2 AMC Examples

The following three examples illustrate the use of each of three possible AMC categories. A suitable AMC category, sample size(s), and the protocol acceptance criteria should be carefully selected to achieve the desired results. The potential post-implementation implications for patient and/or firm resulting from test method replacements should be considered.

6.3.2.1 Demonstrating Noninferiority

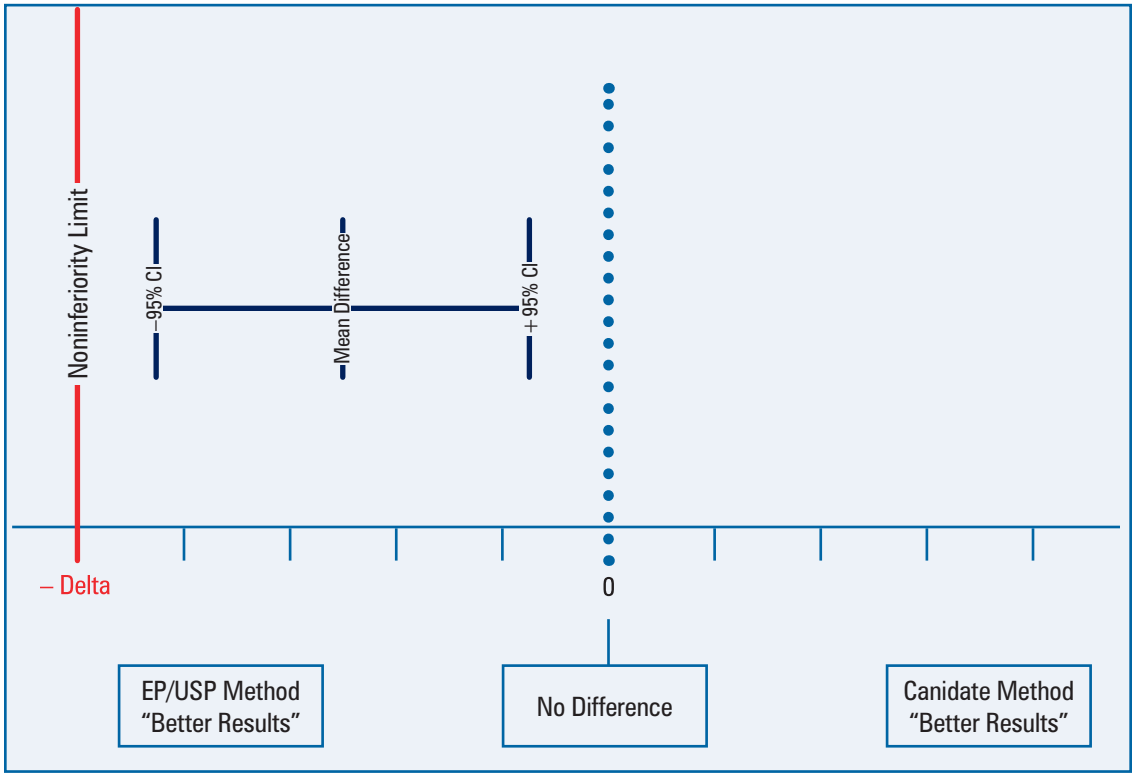
A more rapid and technologically advanced method for sterility testing was validated and compared to the compendial EP/USP Sterility Test (EP 2.6.1/USP <71>). The noninferiority comparison at the 95% confidence level (p=0.05) was chosen with a prespecified Delta of –10% versus the compendial (current) method. A noninferiority test with a Delta of –10% was justified from historical data and the increased sampling plan as part of the validated rapid method use. Noninferiority, equivalence, and superiority are all acceptable outcomes, and the increased testing frequency of daily (n=7 per week) for the new sterility versus twice weekly (n=2 per week) for the EP/USP Sterility test significantly increases the likelihood of detecting organisms with the new method.

The statistical results are given in **Table 6.3.2.1-1** and are illustrated in **Figure 6.3.2.1-1.** The one-sided lower 95% confidence level includes 0 (no difference) and lies entirely to the right of the prespecified Delta of –10%. The comparison results obtained indicate that the candidate method is not inferior to the EP/USP sterility test method at the 95% confidence level.

Table 6.3.2.1-1 Results for the Noninferiority Test: Candidate Method vs. EP/USP Sterility

Method	Positives	Total Samples	Positives-to-Total Samples Proportion
Candidate	225	300	0.75
EP/USP	232	300	0.77
Statistical Results			
Difference = p (new method) –p (EP/USP)			
Estimate for difference: –0.023			
95% lower confidence interval limit for difference: –0.08			
Test for difference = 0 (vs > 0): Z = –0.67 P-Value = 0.75			

Figure 6.3.2.1-1 95% Confidence Interval for Noninferiority Test: Candidate Method vs. EP/USP Sterility



6.3.2.2 Demonstrating Superiority

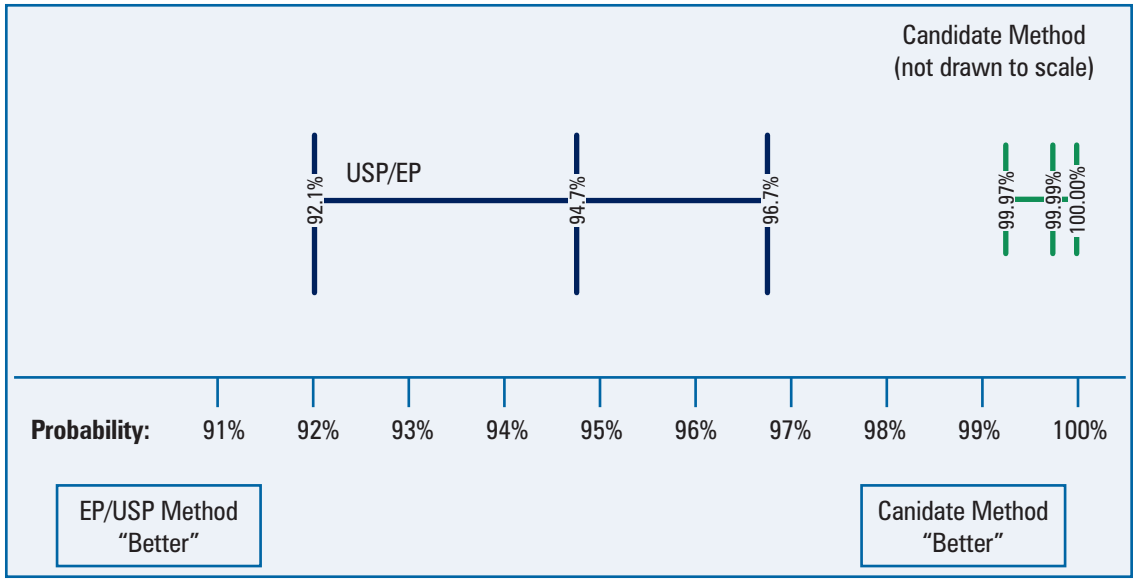
Superiority was demonstrated by comparing the two 95% confidence intervals (EP/USP versus candidate method). When the relative testing frequency of our example under noninferiority of n=7 (candidate method) versus n=2 (EP/USP method) is integrated in our comparison studies, the superiority of the new method could be readily demonstrated. A summary of the statistical results (at 95% confidence) using the data from **Table 6.3.2.1-1** is given in **Table 6.3.2.2-1,** and, the results are graphically illustrated in **Figure 6.3.2.2-1.** The new method's 95% confidence interval (0.9997–1.0000) for the positive-to-fail probability (0.9999) lies entirely to the right of the 95% confidence interval (0.92–0.97) of the compendial method's positive-to-fail probability (0.95).

When the testing frequency was integrated in the AMC study protocol, the superiority test was passed with a much greater relative margin than the noninferiority test. This particular example illustrates the importance of selecting the most-suitable AMC category and acceptance criteria for each situation based upon historical data and the particular use of each method.

Table 6.3.2.2-1 Results for the Superiority Test: New Method (7x per week) vs. EP/USP Sterility (2x per week)

Method	Positives	Total Samples	Probability	95% CI for Probability
Candidate	225	300	0.9999	0.9997 – 1.0000
EP/USP	232	300	0.947	0.921 – 0.967

Figure 6.3.2.2-1 95% Confidence Intervals for Superiority Test: Candidate Method vs. EP/USP Sterility



6.3.2.3 Demonstrating Equivalence

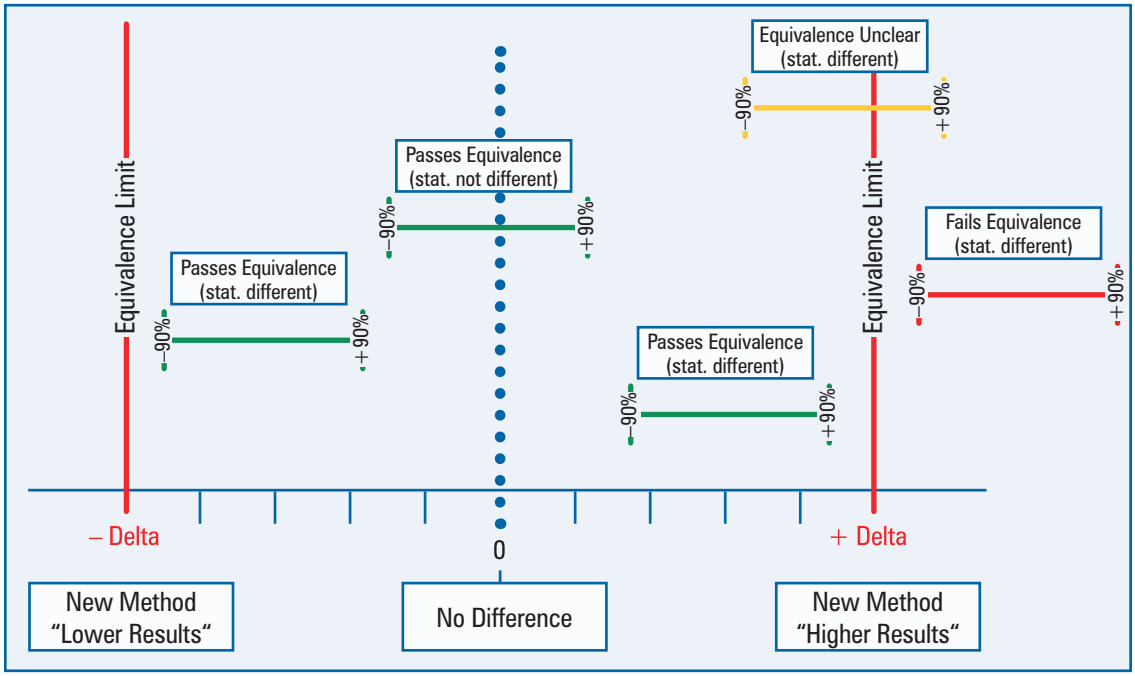
A known impurity is quantitated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) at the final container stage. A validated capillary electrophoresis (CE) method is to replace the SDS-PAGE method. Accuracy and Intermediate Precision should be compared in formal AMC studies. To demonstrate the use of the equivalence model, the accuracy or matching between results from both methods is evaluated below. Data to support the equivalence between both methods should include data from stability studies in addition to release data.

From the analysis of historical final container release and stability data and release specifications (for SDS-PAGE), a Delta of ±1.0% is chosen and justified for the equivalence test between both impurity levels based upon the historical method performance with respect to the specifications (NMT 7.0%). Both methods are run simultaneously (side-by-side) for a total of n=30 final container release and stability samples and results are compared by two-sided paired t-test statistics with a prespecified Delta of ±1.0% (% = reported percent and not relative percent). The paired t-test results are summarized in **Table 6.3.2.3-1** and illustrated in **Figure 6.3.2.3-1.**

Table 6.3.2.3-1 Equivalence Test Results Comparing SDS-PAGE (Reference) to CE

Paired T-Test Results
Hypothesized Difference in Mean: 0%
Minus Delta: –1.0%
Plus Delta: +1.0%
SDS-PAGE Mean (n=30): 3.8%
CE Mean (n=30): 5.1%
Mean difference between CE and SDS-PAGE: 1.2%
90% confidence interval of mean difference : 1.08% – 1.52%

Figure 6.3.2.3-1 90% Confidence Intervals for Equivalence: Candidate Method vs. EP/USP Sterility



The 90% confidence interval of the CE method (1.08% – 1.52%), illustrated in red in **Figure 6.3.2.3-1,** lies entirely on the right side of the positive acceptance limit of 1.0%. The CE results for our impurity are not only statistically higher than those of the SDS-PAGE method, but the expected shift in results is also higher than our prespecified limit.

Possible additional comparison outcomes are illustrated in **Figure 6.3.2.3-1.** If an AMC noninferiority study would have been selected with the left margin –Delta shown all five 90% CI shown to the right of –Delta indicate comparable method performance results. If a superiority test was selected, the three 90% CI to the right of the "No Difference" limit would have yielded passing results. For an equivalence test, only the three 90% CI on the left would have yielded acceptable results. The 90% CI entirely to the right of the +Delta protocol limit indicates non-equivalent or different test method performance. In those cases where the 90% CI overlaps the protocol limit, more AMC samples may have to be tested.

The observed significant difference in results between both methods supports the proportional adjustment of release specifications. This adjustment ensures that an equivalent product quality can be sustained when the CE method is used.

Additional paired assays could be run around the specification limits and/or accelerated stability condition time points to provide assurance that a similar bias between methods still exists at the specification level and that patient safety, product quality, and operational conditions can be sustained after this method change.

7.0 Analytical Method Maintenance

The validation principles of the FDA Guidance on Process Validation can also be applied to AMV practices. Validation should be considered a continuous process ***(5).*** The goal of analytical method maintenance (AMM) is to provide continuous assurance that the validated method remains in a state of control. Therefore, AMM is part of the validation program and should be performed with a similar level of detail und understanding of risks as the preceding AMV studies. AMM can reduce the potential risks associated with changes causing high variation in test results.

7.1 Monitoring Analytical Method Performance

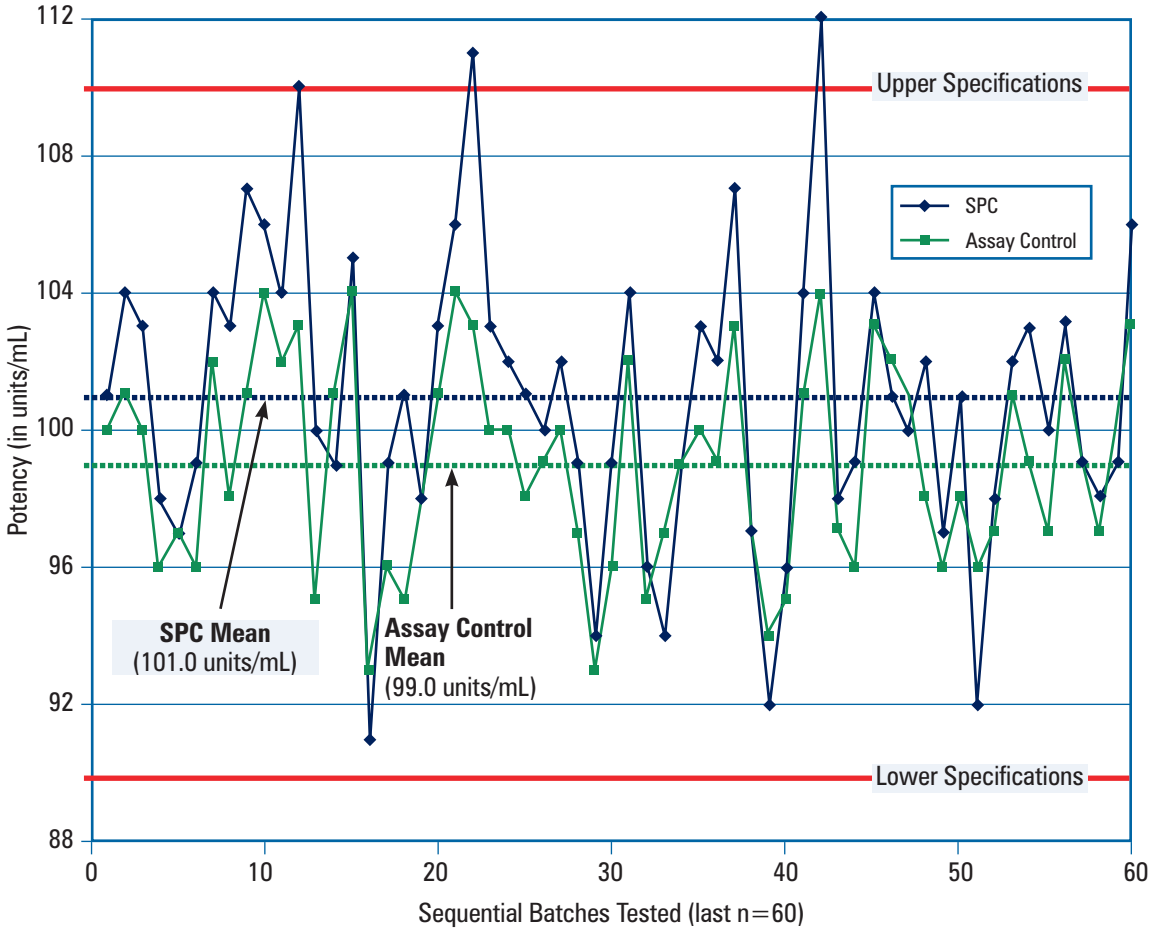
Routine AMM should be monitored using statistical control charts. Similar to process control, AMM control charts are useful to proactively detect and address shift trends, even if they are within accepted control range limits and no invalid test results are generated.

An in-house assay control or reference standard run every time as part of the analytical method procedure are ideal for control charting. Monitoring the control or reference standard values provides an indicator how a test method performs during short and/or long-term time intervals. The observed differences from the expected (mean) reference values are typically indicative how a test method performs with a particular set of conditions (day, analyst, instrument, reagents, etc.).

The laboratory control charts can be combined with the manufacturing process control charts to provide a complete set of manufacturing process control data. By combining the laboratory control data and the manufacturing process control data, an immediate outlier assessment can be performed, if needed. **Figure 7.1-1** illustrates an example of how combining both data sets can provide additional information on statistical process control ***(14, 36).*** Laboratory and/or production process outliers can be immediately identified and addressed using this combined control chart.

In **Figure 7.1-1,** a significant correlation can be observed between both control data sets. In general, a significant and relatively high variation in the laboratory control data (with respect to the production control data) may require a continuous combined monitoring to permit an immediate interference whenever necessary. This undesirable situation may require continuous method performance improvements such as the use of tighter method component qualification criteria or an increase in replicate testing with the intent to lower the apparent variation in laboratory results.

Figure 7.1-1 Combining Laboratory (Assay Control) and Manufacturing Control Charts *(14)*



If the sampling variability (storage, transport, etc.) can also be determined, an estimate of the true process variability can be made similar to **Equation 3.2.1-1.**

In the event an in-house control is not used as part of the method, monitoring other types of control samples, assay calibrators, and/or system suitability criteria using control charting is still providing valuable information about the continuous state of performance. AMV study results and historical data can be used to establish and implement these correlations with the intent to continuously improve the accuracy and reliability of laboratory results. In all cases, the observed changes in the control point(s) used to monitor continuous method performance should significantly correlate with the changes in test results to be suitable for combining laboratory and manufacturing control charts.

7.2 Periodic Review

Historical data should be reviewed periodically by representatives from manufacturing, QC laboratories, and/or other responsible parties to ensure a continuous validated state and a controlled method performance. The method control chart data may provide the most valuable data as any significant method changes can often be traced back using this source. It may not be necessary to review all data sources if the control chart(s) and invalid test result data demonstrate that the test system can be con-

sidered to be in a state of control. Typical data sources and documents that should be reviewed when the method may not be in control are listed below:

- Method history file and/or change control documents
- Combined Historical Production SPC data and Historical Method Performance Charts
- Applicable SOPs
- (Product) specifications
- Historical unexpected results in test samples or method performance indicators
- Test method system suitability specifications
- Test method system suitability failures
- Historical changes in reference standards and/or assay controls
- Historical changes in test method components (reagents)
- Historical analyst performance proficiency/qualification records
- Analytical instrument qualification and calibration data
- Internal and external audit observations
- Relevant corrective and preventative action documents (and their effectiveness)
- Current regulatory guidance documents and other relevant publications

Analytical methods, validated a long time ago, may require formal retrospective verification. A suggested validation status checklist is listed in **Table 7.2-1**. A rigorous review may prevent future unacceptable method performance and the review results may ensure that each test method continues to deliver accurate and reliable test results. It may also ensure that a test method remains compliant and efficient. Changes in production process and/or formulation which may affect the analytical method performance may require some prospective verification studies as described in **Section 4.3.**

Table 7.2-1 Suggested Checklist Items to Assess Validation Status ***(14, 36)***

AMV and Method Performance Checklist Items	Results	Comments
Test Method Number/Title/Revision:		
Process Step/Product Sampling Point(s):		
Most Recent Validation/Verification Date:		
Specifications and/or Action Levels Supported:		
ICH Q2(R1) Test Method Category:		
Suitable Accuracy Demonstrated in AMV?		
Suitable Repeatability Precision Demonstrated in AMV?		
Suitable Intermediate Precision Demonstrated in AMV?		
Suitable Specificity Demonstrated in AMV?		
Suitable Linearity Demonstrated in AMV?		
Suitable Assay Range Demonstrated in AMV?		
Suitable Detection Limit Demonstrated in AMV?		
Suitable Quantitation Limit Demonstrated in AMV?		
Suitable Robustness Demonstrated in AMD/AMV?		
Suitable System Suitability Demonstrated in AMV?		
Number of Valid Test Runs Over Last 12 Months		
Number of Invalid Test Runs Over Last 12 Months		
Calculate Invalid Rate/Percentage:		
Statistical Assay Control Limits (ex., 3 Standard Deviations):		
Test System in Control?		
Changes to Test System After AMV: If yes, provide more information:		
Most Recent AMV Studies Acceptable? If no, provide risk-based priority for revalidation for VMP:		
Method Performance Acceptable? If no, provide risk-based priority for method improvement list:		
QC Signature:		
QA Signature:		

7.3 Replacing Analytical Method Components

Over time it may become necessary to replace or add components critical to the analytical method (for example, a new instrument). Risks should be assessed following similar concepts as presented in **Section 3.1.**

Historical data sources such as AMV results for Intermediate Precision, control charts, and/or AMD/AMQ results for the method's Robustness should be reviewed to assess the potential impact to post-implementation method performance. Equivalency between the validated component and its replacement should be demonstrated with actual laboratory data prior to the implementation of replacements of critical components. Similar to method replacements (**Section 6**) but typically with less data pairs compared, equivalency for the changed component may be demonstrated by lack of bias as well as comparable reliability for all quantitative methods. Other comparative characteristics may apply based on the use of the method. A noninferiority type comparison may be used for qualitative methods.

A more extensive qualification approach is expected when replacing a reference standard. Future data continuity should be ensured through parallel testing of the new and old reference standard with the intent to sustain test data continuity during and after this change process.

8.0 AMV Discrepancies/Failures

A validation failure is the result of failing to pass protocol acceptance criteria (preset test method performance specifications). The setting of risk-based AMV protocol limits based on risks to patient and/or firm was extensively discussed in **Section 3.2.** When a validation failure occurs, a root cause analysis and an impact assessment should be conducted and documented through investigations of validation failures or unexpected results.

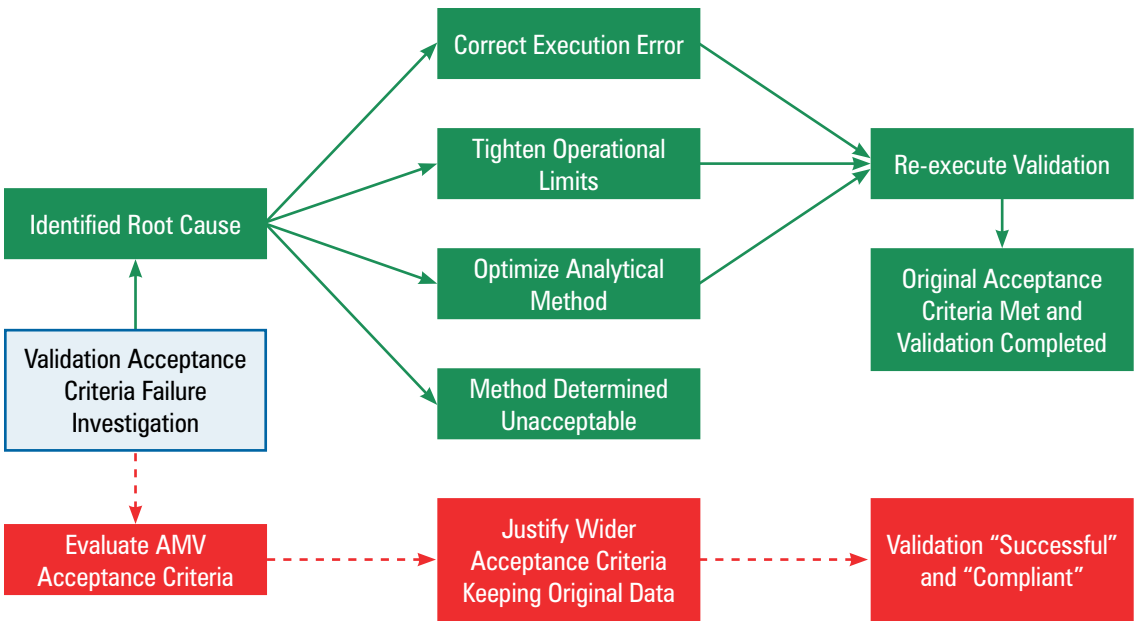
Figure 8.0-1 illustrates the management of an AMV failure and its recovery process. Once a validation failure is observed, the investigation should be documented in a formal report. Within the upper pathway, the first choice in **Figure 8.0-1** is to re-execute with the current protocol acceptance criteria based on having found a likely root cause. Correcting an unexpected error will not change or improve anything for the routine test method performance. For example, inappropriate reference material was used and the validation study is re-executed using appropriate material. This could be justified simply based on the fact that the original acceptance criteria for the test method validation remain unchanged.

The second choice, "Tighter Operational Limits", would require running the test method system under more stringent operational parameters (e.g., reduce tolerance for incubation time, increase mixing time). Although these limitations may indicate that the test method may not be robust, this adjustment may lead to improved intermediate precision results.

The third choice, choosing to optimize the analytical method, may have the greatest effect on the future test method performance. A rigorous method optimization effort may often result in the most noticeable increase in test method performance. Because risk, scope, budget, and timelines are vital for many firms, all aspects that may impact patient safety, process or product quality, compliance/inspections, completion time, costs, chances for improvement success, and short- and long-term benefits should be balanced.

Choosing the lower pathway, "Evaluate AMV Acceptance Criteria", resulting in the lack of root cause identification and/or re-validation depending on the investigation outcome, should be avoided and is therefore colored red in **Fig. 8.0-1.** If the risk assessment(s) and setting of acceptance criteria, as suggested in **Section 3.0,** are followed, proceeding within the lower pathway may not be justifiable.

Figure 8.0-1 Failing Acceptance Criteria - The "Recovery Mission" ***(14)***



8.1 Investigation and Decision Process

As part of the acceptance criteria failure investigation, illustrated in **Figure 8.0-1,** evaluation of answers to a list of common questions will lead to the most suitable path forward. **Table 8.1-1** lists typical questions that may be considered to help assess potential risk to patients, compliance, project completion, and the firm ***(14).*** By assessing the risks from all suggested information sources, the level of the effort and time required for AMV completion can be estimated. The release of unsuitable material or product through the use of unreliable and/or inaccurate analytical methods constitutes a risk to patients, whereas, an out-of-specification (OOS) result caused by an inaccurate analytical test result may be considered a risk to the firm.

Questions 1-4 in **Table 8.1-1** are an impact assessment addressing safety, quality and efficacy identifying potential risk primarily to patients, although they also support assessment of risk to the firm. The overall outcome of this set of questions should help drive the decision regarding the aspects (upper loop) of the method that needs to improve. Questions 3-4 lead to a better understanding of the historical test method performance that may not have been sufficiently known or captured in the AMD and/or AMQ report.

- **Question 1:** Using probability calculations, the criticality of a particular method performance characteristic that failed the protocol acceptance criteria can be assessed with respect to the probability of obtaining test results within or outside of the corresponding specification(s).
- **Question 2:** The criticality of the test result is assessed with respect to overall product safety, efficacy, and quality.
- **Question 3:** The reliability of the test method is assessed from the AMD and/or AMQ data and related documentation, if applicable and/or available, such as the reference standard assignment, AIQ, AMT, and AMV (if previously validated).
- **Question 4:** To complete the investigation process, supporting documentation such as laboratory notebooks of the AMD scientist should be thoroughly reviewed and interviews with relevant personnel should be conducted and documented.

Questions 5-7 are intended to assess the overall risk(s) to the firm's compliance standing and the outcome of future regulatory inspections by reviewing recent regulatory audit notes and observations. Failures and potential corrective actions may not be sufficiently discussed in regulatory submissions. An overall compliance gap analysis with respect to the occurred AMV failure(s) may therefore suggest particular corrective and/or preventive actions (CAPA) that may fit best the overall need.

- **Question 5:** An objective comparison between relevant passing and failing AMVs may provide a measurable level of certainty as to how high the overall compliance and inspection risk will be depending on the possible actions taken.
- **Question 6:** Assuming that a standard procedure for the investigation process of AMV failures does not exist at the time of the failure, the impact of this failure to the overall firm's validation process and compliance should be assessed. The post-failure CAPA process may include the development of a standard procedure to address future AMV failures.
- **Question 7:** The overall investigation may include the review of regulatory expectations noted during recent audits, because regulatory expectations for validation and the handling of failures can differ significantly among regulatory agencies and the type of drug product and/or process affected.

Question 8 integrates the project completion components into the overall impact assessment and management process. Management should be sufficiently informed and involved to support the proper allocation of resources and time to complete the overall failure resolution process.

- **Question 8:** Project completion can be a significant factor for a firm at the time of the failure occurrence. The time and costs needed to complete the identified method performance improvements may be openly integrated into the overall decision process.

Table 8.1-1 Checklist of Most Common Questions and Possible Information Sources

Question No.	Examples of Questions	Possible Information Source(s)
1	*Did we fail to pass a critical protocol acceptance criterion (or several) such as intermediate precision when high variability could cause OOS results?*	*Check for criticality and corresponding likelihood for OOSs to occur.*
2	*Are results generated by this test method critical to assess product safety or product/process quality, or efficacy?*	*Consider production process stage, and impact to safety, quality or efficacy.*
3	*Did we have previous AMV failures (and discrepancies) with this test method?*	*If this is not a new method, review previous AMV(s).*
4	*Were there any (failing) data sets generated during AMD that were not discussed in the AMD report?*	*Review laboratory notebooks from AMD scientists and (if necessary) conduct interviews with AMD scientists.*
5	*Has this kind of failure occurred before and how did we handle this?*	*Count failures versus successful completions and review previous recovery processes.*
6	*Were there previous inspection observations for validation processes and/or failures?*	*Review previous regulatory and internal observation notes.*
7	*What is our current overall compliance standing with regulatory agencies?*	*Review previous commitments of firm and current regulatory expectations.*
8	*What is the impact to this project and connected projects?*	*Review and discuss project timelines and cost with project managers.*

9.0 References

1. Quality Guideline Q2(R1): Validation of Analytical Procedures: Text and Methodology; International Conference on Harmonisation: 2005. www.ich.org.
2. Quality Guideline Q8(R2): Pharmaceutical Development; International Conference on Harmonisation: 2009. www.ich.org.
3. Quality Guideline Q9: Quality Risk Management; International Conference on Harmonisation: 2005. www.ich.org.
4. Quality Guideline Q10: *Pharmaceutical Quality System*; International Conference on Harmonisation: 2008. www.ich.org.
5. *Guidance for Industry: Process Validation: General Principles and Practices,* U.S. Food and Drug Administration: 2011. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070336.pdf (accessed July 27, 2012)
6. Quality Guideline Q6A: *Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances;* International Conference on Harmonisation: 1999. www.ich.org.
7. *Guidance for Industry: Bioanalytical Method Validation*, U.S. Food and Drug Administration: 2001.
8. Quality Guideline Q6B: *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products Q6B;* International Conference on Harmonisation: 1999. www.ich.org.
9. Montgomery, D. *Design and Analysis of Experiments*, 6th Ed.; John Wiley & Sons: Hoboken, NJ, 2005.
10. Efficacy Guideline E9: *Statistical Principles for Clinical Trials*; International Conference on Harmonisation: 1998. www.ich.org.
11. Quality Guideline Q5E: *Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process;* International Conference on Harmonisation: 2005. www.ich.org.
12. *Guidance for Industry, Investigating Out-of-Specification (OOS) Test Results for Pharmaceutical Production;* U.S. Food and Drug Administration: 2006.
13. Quality Guideline *Q7A: Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients;* International Conference on Harmonisation: 2011. www.ich.org.
14. Krause, S.O. *Validation of Analytical Methods for Biopharmaceuticals: A Guide to Risk-Based Validation and Implementation Strategies*, PDA/Davis Healthcare International Publishing: River Grove, IL, 2007.
15. Pryseley, A., et al. *Estimating Precision, Repeatability, and Reproducibility From Gaussian and Non-Gaussian Data: A Mixed Models Approach. J Appl. Stat.* **2010**, *10,* 1729-1747.
16. ASTM E29-02, *Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications;* ASTM International: West Conshohocken, PA, 2002.
17. ASTM E1488-02, *Standard Guide for Statistical Procedures to Use in Developing and Applying ASTM Test Methods;* ASTM International: West Conshohocken, PA, 2002.
18. ASTM D4854-95, *Standard Guide for Estimating the Magnitude of Variability from Expected Sources in Sampling Plans;* ASTM International: West Conshohocken, PA, 2001.
19. ASTM E1169-02, *Standard Guide for Conducting Ruggedness Tests;* ASTM International: West Conshohocken, PA, 2002.
20. ASTM D4853-97, *Standard Guide for Reducing Test Variability;* ASTM International: West Conshohocken, PA, USA, 2002.
21. General Chapter <1225> *Validation of Compendial Methods. USP 35-NF 30,* United States Pharmacopeia: www.usp.org.
22. Quality Guideline Q1A (R2): *Stability Testing of New Drug Substances*; International Conference on Harmonisation: www.ich.org, 2003.
23. Quality Guideline Q5C: *Stability Testing of Biotechnological/Biologic Products*; International Conference on Harmonisation: www.ich.org, 1995.
24. Krause, S. *Good Analytical Method Validation Practice, Part 1: Setting-Up for Compliance and Efficiency.* J. Val. Technol. **2002**.
25. ASTM E456-08e3, *Standard Terminology Relating to Quality and Statistics;* ASTM International: West Conshohocken, PA, USA, 1996.
26. General Chapter <1226> *Verification of Compendial Procedures. USP 35-NF 30,* United State Pharmacopeia: www.usp.org.
27. General Chapter <1224> *Transfer of Analytical Procedures. USP 35-NF 30,* United State Pharmacopeia: www.usp.org.
28. *ISPE Good Practice Guide: Technology Transfer;* International Society for Pharmaceutical Engineering: Tampa, FL, 2003.
29. Proposed General Chapter: <1033> *Biological Assay Validation. Pharmacopoeial Forum.* **2009,** *35*(2), *349-367.*
30. Krause, S. *Qualifying Release Laboratories in Europe and the United States, BioPharm Internat.* **2004.**

31. *Guidance for Industry, Statistical Approaches to Establishing Bioequivalence;* U.S. Food and Drug Administration: 2001.

32. Barker, L. et al. *Equivalence Testing for Binomial Random Variables: Which Test to Use?*. Am. Stat. **2001,** *55,* 279-87.

33. United States Pharmacopeia, 35-National Formulary 30, First Supplement, General Chapters: <1010>, *Analytical Data – Interpretation and Treatment*, May 2012.

34. Miller, J.N.; Miller, J.C. *Statistics and Chemometrics for Analytical Chemistry* 4th ed.; Prentice Hall: Harlow, England, 2000.

35. Krause, S. *Submitting Advanced Bioanalytical Test Methods for Regulatory Approval.* BioPharm Internat. **2005.**

36. Krause, S. *Analytical Method Maintenance.* PharmTech Europe. **2006.**

Terms of Usage

An **Authorized User** of the electronic PDA Technical Report is a PDA Member in good standing or the purchaser of the Technical Report.

Authorized Users are permitted to do the following:

- Search and view the content of the PDA Technical Report
- Download the PDA Technical Report for the individual use of an Authorized User
- Print the PDA Technical Report for the individual use of an Authorized User
- Make a reasonable number of photocopies of a printed PDA Technical Report for the individual use of an Authorized User

Authorized Users are not permitted to do the following:

- Allow anyone other than an Authorized User to use or access the PDA Technical Report
- Display or otherwise make any information from the PDA Technical Report available to anyone other than an Authorized User
- Post articles from the PDA Technical Report on Web sites, either available on the Internet or an Intranet, or in any form of online publications without a license from PDA, Inc.
- Transmit electronically, via e-mail or any other file transfer protocols, any portion of the PDA Technical Report
- Create a searchable archive of any portion of the PDA Technical Report
- Use robots or intelligent agents to access, search and/or systematically download any portion of the PDA Technical Report
- Sell, re-sell, rent, lease, license, sublicense, assign or otherwise transfer the use of the PDA Technical Report or its content
- Use or copy the PDA Technical Report for document delivery, fee-for-service use, or bulk reproduction or distribution of materials in any form, or any substantially similar commercial purpose
- Alter, modify, repackage or adapt any portion of the PDA Technical Report
- Make any edits or derivative works with respect to any portion of the PDA Technical Report including any text or graphics
- Delete or remove in any form or format, including on a printed article or photocopy, any copyright information or notice contained in the PDA Technical Report
- Combine any portion of the PDA Technical Report with any other material

To License or purchase Reprints or Hardcopies

Licensing: Site licenses and licenses to distribute PDA Technical Reports can be obtained for a fee. To learn more about licensing options and rates, please contact: Janny Chua, Publications Manager, +1 (301) 656-5900, ext. 133. Written correspondence can be sent to: PDA, Inc. Attn: Janny Chua, 4350 East West Highway, Suite 150, Bethesda, MD 20814.

Reprints: Reprints of PDA Technical Reports can be purchased for a fee. To order reprints, please contact: Janny Chua, Publications Manager, +1 (301) 656-5900, ext. 133. Written correspondence can be sent to: PDA, Inc. Attn: Janny Chua, 4350 East West Highway, Suite 150, Bethesda, MD 20814.

Hardcopies: Hardcopies of PDA Technical Reports are available on demand for a fee. The PDA membership benefits include discounted pricing for printed copies of PDA Technical Reports. Go to www.pda.org/bookstore to order a hardcopy. PDA members will need their PDA identification number and password to receive the discount.

About PDA Technical Reports

PDA Technical Reports are global consensus documents, prepared by member-driven Task Forces (listed on inside front cover) comprised of content experts, including scientists and engineers working in the pharmaceutical/biopharmaceutical industry, regulatory authorities and academia. While in development, PDA Technical Reports are subjected to a global review of PDA members and other topic-specific experts, often including regulatory officials. Comments from the global review are then considered by the authoring Task Force during the preparation of the final working draft. The level of expertise of the Task Force and those participating in the global review ensure a broad perspective reflecting best thinking and practices currently available.

The final working draft is next reviewed by the PDA Advisory Board or Boards that sanctioned the project. PDA's Advisory Boards are: Science Advisory Board, Biotechnology Advisory Board, Regulatory Affairs and Quality Committee, and Audit Guidance Advisory Board. Following this stage of review, the PDA Board of Directors conducts the final review and determines whether to publish or not publish the document as an official PDA Technical Report.

Bethesda Towers
4350 East West Highway
Suite 200
Bethesda, MD 20814 USA
Tel: +1 (301) 656-5900
Fax: +1 (301) 986-0296
E-mail: info@pda.org
Web site: www.pda.org