The AAPS Journal (© 2014)
DOI: 10.1208/s12248-014-9696-2

White Paper

Workshop Report: Crystal City V—Quantitative Bioanalytical Method Validation and Implementation: The 2013 Revised FDA Guidance

Brian Booth,[1] Mark E. Arnold,[2,12] Binodh DeSilva,[2] Lakshmi Amaravadi,[3] Sherri Dudal,[4] Eric Fluhler,[5] Boris Gorovits,[6] Sam H. Haidar,[1] John Kadavil,[1] Steve Lowes,[7] Robert Nicholson,[8] Marie Rock,[9] Michael Skelly,[1] Lauren Stevenson,[3] Sriram Subramaniam,[1] Russell Weiner,[10] and Eric Woolf[11]

Received 26 September 2014; accepted 31 October 2014

Abstract. In September 2013, the FDA released a draft revision of the Bioanalytical Method Validation (BMV) Guidance, which included a number of changes to the expectations for bioanalysis, most notably the inclusion of biomarker assays and data. To provide a forum for an open, inclusive discussion of the revised draft BMV Guidance, the AAPS and FDA once again collaborated to convene a two-and-a-half day workshop during early December 2013 in Baltimore, MD, USA. The resulting format embodied extensive open discussion and each thematic session included only brief, concise descriptions by Agency and industry representatives prior to opening the floor discussion. The Workshop was built around four thematic sessions (Common Topics, Chromatographic, Ligand-Binding Assays, and Biomarkers) and a final session with international regulators, concluding with a review of the outcomes and recommendations from the thematic sessions. This Workshop report summarizes the outcomes and includes topics of agreement, those where the FDA will consider the Industry's perspective, and those where the workshop provided a first open dialogue. This article will be available to the bioanalytical community at http://www.aaps.org/BMV13.

KEY WORDS: bioanalytical method validation; Crystal City V; FDA guidance.

INTRODUCTION

The quantitative measurements of drugs, metabolites, and biomarkers in nonclinical and clinical studies provide essential information in the assessment of safety and efficacy of drugs. Drug or biomarker concentrations frequently serve as the primary or secondary endpoints of many clinical studies in drug development. Consequently, the reliability or quality of that data underpins the study outcome. For example, in bioequivalence studies, the pharmacokinetic (PK) comparison is the basis for approval, thus the quality of the concentration data is essential. Similarly, as biomarkers and the resulting pharmacodynamic (PD) interpretations are exploited more extensively in establishing efficacy and labeling claims, the quality of these determinations will require greater demonstration of the assay quality and usage. In all cases, these PK or PD measurements are based on established principles and scientists can utilize a common, vetted paradigm of practices, independent of the analytical platform to demonstrate that the assays provide reliable data.

The evolution of this bioanalytical paradigm began with the first American Association of Pharmaceutical Scientists (AAPS)/Food and Drug Administration (FDA) Bioanalytical Workshop in 1990 (1). Scientists in the bioanalytical field worked with the regulatory community to establish a common language and expectations in generating pharmacokinetic data for drugs and metabolites. These validation principals were introduced into regulations by Health Canada in 1992 (2) and then by the FDA which published the first edition of its Guidance on Bioanalytical Method Validation (BMV) (3) in 2001. Since then, the dialogue has broadened significantly through scientific conferences not only within the USA, but globally. The last decade has seen the additional introduction of BMV regulation in Brazil (4), the EU (5), and Japan (6), with

[1] U.S. Food and Drug Administration, Silver Spring, MD, USA.
[2] Bristol-Myers Squibb Co., Princeton, NJ, USA.
[3] Biogen Idec, Cambridge, MA, USA.
[4] Roche Innovation Center, Basel, Switzerland.
[5] Pfizer Inc., Pearl River, NY, USA.
[6] Pfizer Inc., Andover, MA, USA.
[7] Quintiles Bioanalytical and ADME Labs, Ithaca, NY, USA.
[8] PPD, Richmond, VA, USA.
[9] WIL Research, Skokie, IL, USA.
[10] Merck Research Laboratories, Rahway, NJ, USA.
[11] Merck Research Laboratories, West Point, PA, USA.
[12] To whom correspondence should be addressed. (e-mail: mark.arnold@bms.com)

ABBREVIATIONS: A&P, Accuracy and Precision; ADC, Antibody-Drug Conjugate; BQL, Below Quantitation Limit, <LLOQ is also used; Conjugated Antibody, Antibody with DAR equal or greater than 1; DAR Drug Antibody Ratio; Dx, Diagnostic; HQC, High Quality Control; LLOQ, Lower Limit of Quantitation; LQC, Low Quality Control; MQC, Mid Quality Control; ULOQ, Upper Limit of Quantitation.
DEFINITIONS: Unconjugated Drug, drug spontaneously released in vivo from an ADC; Antibody-Conjugated Drug, drug conjugated to the antibody moiety; Total Antibody, Antibody with DAR equal or greater than 0. Includes conjugated and fully unconjugated antibody; Mock Clinical Sample, Samples prepared by pooling samples or spiking over endogenous levels. Used in the context of demonstrating biomarker stability on storage and measured periodically with a Dx.

Published online: 31 December 2014

other countries considering new regulation or applying existing regulations from other regions. In September 2013, the FDA released a draft revision (7) of the BMV guidance, and the discussion continued. The current draft document included a number of changes to the expectations for bioanalysis, most notably the inclusion of biomarker assays and data.

To provide a forum for an open, inclusive discussion of the revised draft BMV Guidance, the AAPS and FDA once again collaborated to convene a two-and-a-half day workshop during early December 2013 in Baltimore, MD, USA. Although not held in Arlington, VA, Crystal City V (CCV) built upon its historic precedents to facilitate an open dialogue between the industry and Agency. An objective set forth early in the planning was the need for more extensive open comment sessions. The resulting format embodied that concept and each thematic session included only brief, concise descriptions by Agency and industry representatives prior to opening the floor discussion. Those introductory presentations focused on what had changed and why, and it highlighted areas of industry concern to be discussed in the open sessions. Each open session was moderated by Agency and industry representatives, included a panel of subject matter experts (SME) from the FDA and industry, and utilized additional SMEs from the FDA and industry to roam among the audience and stimulate the discussion.

The Workshop was built around four thematic sessions (Common Topics, Chromatographic, Ligand Binding Assays, and Biomarkers) and a final session with international regulators, concluding with a review of the outcomes and recommendations from the thematic sessions. All of these sessions highlighted the changes in the pharmaceutical industry, in particular the growth of biotechnology-based therapies. Over 450 scientists attended the Workshop, with 25% of participants coming from 18 countries outside the USA. In many cases, the attendees represented not only their own companies but also the numerous regional and global bioanalytical consortia (*e.g.*, Global Bioanalysis Consortium, International Consortium on Innovation & Quality in Pharmaceutical Development, European Bioanalysis Forum, Japanese Bioanalysis Forum, Applied Pharmaceutical Analysis—India, AcBio, Canadian Forum for Analytical and Bioanalytical Sciences, Chinese Bioanalysis Forum, Global CRO Council), thus bringing the concerns, thoughts, and interest of an even larger bioanalytical community into the discussion.

As with previous Crystal City meetings, part of the objective was to discuss changes in regulatory thinking and the science applied to new drug development that had occurred since the last Workshop. This Workshop included discussions on such science-driven topics such as the status of incurred sample reanalysis (ISR) after several years of application, new immunoassay technology, antibody-drug conjugates (ADCs), and the application of LC-MS/MS for protein quantitation. Additionally, the Agency through the guidance and its presentations highlighted areas of additional focus for chromatographic and ligand-binding assays, such as stock solution expiry and most notably, the inclusion of biomarker assays within the Guidance. Extensive discussions, from multiple points of view were generated around the topics presented.

The salient outcomes of CCV can be broken down into two major categories: topics for which there was general consensus between industry and FDA based on the clarifications and discussions which took place and topics for which no consensus was achieved between industry and the Agency. The issues for which no consensus was reached were characterized by two situations: topics on which viewpoints differed and topics in which the science is still too new. In cases where viewpoints differed, the Agency agreed to consider industry positions within their deliberations leading to the final guidance. In the latter case, the Agency sought the current thinking and experience of industry scientists, which will help guide future recommendations in these areas. The following provides a summary of all of the discussions which took place at CCV delineated by these categories.

CHROMATOGRAPHIC ASSAYS: CONSENSUS TOPIC

The basic expectations for BMV of chromatographic assays have been previously defined in the 2001 FDA guidance (3) and subsequent conference reports on the topic (8,9).

METHOD VALIDATION

Several aspects of *cross* and *partial validation* received extensive discussion. It was not possible to further refine the criteria for either type of validation due to the breadth of the issue. However, further clarification was added to the following issues:

Partial Validation

The need to validate a new analyst brought into a study was discussed. The consensus that was reached is that new analysts are expected to be trained/qualified on relevant techniques prior to supporting studies. There is no expectation that training or qualification data should be included in a validation report, although it must be retained in training or study files. Performing a single test run that passes acceptance criteria was suggested as an example for a new analyst.

Regarding the need to partially validate changes of an anticoagulant counter-ion, the consensus view was that validation is not required unless an effect is observed during the method development.

Cross Validation

Due to varied experiences and the wide range of scenarios, no further detail regarding acceptance criteria were elucidated, but it was agreed that criteria should be defined in a "standard operating procedure" (SOP) or validation plan document *a priori*. Furthermore, it was recommended that cross validation analyses should be performed in both laboratories (or between methods), using shared spiked matrix quality control (QC) samples and non-pooled incurred subject samples. Pooled incurred samples can be used when insufficient volume exists.

Reference Standards

Considerable discussion of Section III A of the draft Guidance (7) centered on use of reference standards in chromatographic methods. The section clarifies types of reference standards in terms of their source and continues to designate expectations around supporting documentation including expiration date, certificates of analysis (COAs) and evidence of purity. The extension of reference standard documentation to include internal standards (IS) was also

presented. There was a good discussion at the Workshop on the proposed documentation for the internal standard material. Industry representatives noted that it is not always possible to obtain a formal CoA for stable label (SLIS) or analog internal standards due to the limited amount of material synthesized. It was also suggested that purity and stability of the IS is generally not critical for typical chromatographic assays (*e.g.*, LC-MS/MS) as long as the suitability can be demonstrated (*e.g.*, no interference from the IS to the analyte(s) of interest). In response, the Agency clarified that CoAs for the IS are preferred, but other documentation characterizing the purity is acceptable when a formal CoA is not available. In any case, the suitability of the material for its intended use should be demonstrated prior to and/or during use. This extended to an agreement that the stability of the IS need not be demonstrated as long as suitability (lack of interference) is demonstrated.

Matrix Effects and Selectivity

At the current Workshop, clarification was sought on the types of matrix effects that should be studied during validation. The goal of the matrix effect assessment is to demonstrate consistent assay performance across multiple sources of matrix that reflect the anticipated population from which samples are expected. Variations of lipid and specific or total protein concentrations which may result from the disease indication should be addressed. The impact of hemolyzed and lipemic samples may be assessed during method development; a specific assessment of these factors is not expected during most validations, unless the assay is intended to support studies where a high percentage of samples may be hemolyzed or are from hyperlipemic subjects.

While validation experiments should assess ion suppression/enhancement (matrix effect), calculation of matrix factor (MF) is not an expected component of validation, although it is understood that proper investigation of ion suppression/enhancement provides data necessary to make the MF calculation if applicable to the investigation. MFs can be calculated for both the analyte and internal standard in a given matrix. Ideally, the variation in MFs for the analyte is compensated for, at least to some degree, by the variation of the MF of the internal standard. Stable label internal standards typically provide the best degree of compensation, hence they are recommended for MS-based assays.

Selectivity is the ability of an analytical method to differentiate and quantify the analyte of interest in the presence of matrix components in the sample. Assay selectivity assessment is an important component of validation. The consensus at the workshop was that selectivity should be assessed in six individual lots of matrix. These lots may include hemolyzed or hyperlipemic samples as deemed necessary. Generally, gender assessments of selectivity are not expected, but consideration should be given to any unique attributes of the analyte that could be impacted by gender. If it is anticipated that samples from specialized study populations (*e.g.*, renally impaired) may be analyzed with an assay, consideration should be given to assessing selectivity in blank matrix from the target population or control samples obtained from such subjects.

Validation Run Acceptance

Feedback on the draft Guidance indicated a level of confusion around what is proposed regarding the anticipated number of validation runs that encompass a bioanalytical method validation. This was reflected in the Workshop discussions with specific focus on the reference to the total number of runs ("minimum of six runs conducted over several days") and the number of runs required to demonstrate within-run and between-run accuracy and precision (A&P). A need emerged during the discussion to address a fundamental concern that scientists, auditors, and inspectors share. That is, the potential of "cherry-picking" only validation runs that meet *a priori* acceptance criteria does not provide an accurate representation of method performance. From the discussion, a consensus was derived to address a better estimate of the A&P of a bioanalytical method during validation than currently employed.

It was proposed that method A&P should be determined from QC performance independent of *a priori* criteria. To accommodate this, run acceptance during A&P validation runs should be based on calibration curve performance only. Back calculated calibrator concentrations should meet acceptance criteria (75% of calibrators within ±15% of nominal, except at the lower limit of quantitation of the assay (LLOQ) where they should be within ±20%) to pass A&P validation runs only. Using a minimum of three runs, the A&P statistics should then be calculated from the QCs using at least five replicates in accepted runs at four concentrations including the LLOQ, as well as low (LQC), middle (MQC), and high (HQC) concentrations, typically corresponding to those used for run acceptance QCs. A specific assessment of A&P at the upper limit of quantitation of the assay (ULOQ) is not required for chromatographic methods. Furthermore, at least one A&P run should use calibrators and validation samples prepared from independent stock solutions. For all other validation runs, both calibrators and samples may be prepared using a single verified stock solution. The A&P results for the QCs in all passing runs should be subsequently reported. To reiterate, this approach avoids including QC criteria in run acceptance in A&P validation runs and delineates A&P *performance* as the totality of QC performance is distinct from the calibration curve performance that determine whether a run passes or fails.

For other validation runs (*i.e.*, those not associated with A&P and typically with stability assessments), run acceptance should be based on calibration standards (as defined above) and the performance of run acceptance QCs (*i.e.*, analytical QCs) consistent with current regulatory guidance. Specifically, at least three concentrations of QCs in at least duplicate should be incorporated into each run (LQC, MQC, and HQC). For the run to pass, 75% of calibrators should be within ±15% of nominal, except at the LLOQ where they should be within ±20% and at least 67% (*e.g.*, at least four out of six) of all QCs should fall within ±15% of their respective nominal concentrations and at least 50% of the QCs at each level should pass.

Recovery

Based on the description provided in the draft guidance, a clarification was sought on the appropriate approach for determining recovery during validation. The consensus view was that recovery, as applied to chromatography-based

bioanalytical methods, should be defined as a comparison of the instrumental response associated with extracted samples compared to that of extracts of control matrix spiked with analyte post extraction. The response of the samples spiked post extraction represents 100% recovery. Recovery of the analyte need not be 100%, but the extent of recovery of an analyte and of the internal standard should be consistent and reproducible. Recovery experiments should be performed by comparing the analytical results for extracted samples at three concentrations (low, medium, and high) with those of corresponding spiked control extracts.

Method Use

A number of topics were discussed related to the application of validated methods during study sample analysis. For example, the Agency clarified its concern that equipment conditioning and system suitability testing with calibrators, QCs, and study samples from a pending run could risk biasing the decisions on acceptance and rejection of the pending run, or of the study sample results. There was no concern about using separately prepared extracts or reusing extracts from a completed run for purposes of conditioning or system suitability.

The draft guidance recommends adding new QCs at different concentrations, if the original QC concentrations were not representative of study sample results from early analyses. The consensus was that partial validation of new calibration and QC concentrations would be appropriate only if the new concentrations were not bracketed by earlier validations. Ordinarily, demonstration of the prepared concentration would suffice when adding QC concentrations in validated regions of the calibration range for a better demonstration of assay performance near the most critical concentrations of study samples.

During sample analysis, when QCs fail during individual runs that meet acceptance, the draft guidance recommends reporting these values. The Agency indicated that it is interested in evaluating variability in performance, including any outlier values. Where results are generated for QCs in failing runs, documentation of their results in the report is recommended but it is not necessary to include those QCs in the accuracy and precision statistical analysis (see later discussion on Documenting and Reporting).

During most bioequivalence studies, and during other studies designed with repeated measures from individual subjects, it is desirable to minimize the between-run variability. The draft guidance recommends that samples from a given subject should be analyzed in a single run. The consensus was that this should be the practice when practicable. Products such as those therapeutic proteins with extremely long elimination half-lives was an example discussed where this approach is not pragmatic. Whatever approach is taken for sample analysis, it should include consideration and steps to minimize variability between periods.

CHROMATOGRAPHIC ASSAYS: NON-CONSENSUS TOPIC

This section describes the topics for which there was lack of agreement at the current AAPS/FDA Bioanalytical Method Validation Workshop and is presented to provide the reasons for no consensus, and promote further discussion.

Analyte Stability in Presence of Concomitant Medications

A majority of Workshop attendees maintained that historical data suggest that concomitant medications pose little risk for influencing instability of analytes, and, therefore, assessing the stability of analytes in the presence of concomitant medications adds little value. However, there was some discussion of rare cases when instability was observed with other analyte(s) but this was attributed to chemistry issues, and only in such cases should additional stability be conducted. The Agency's argument for conducting additional stability evaluations with concomitant medications, particularly fixed-dose combinations, was that this is an acceptable practice to assure stability in the known presence of two (or more) analytes, since stability is an integral part of method validation. Some conference attendees indicated that the analytical sites may not always be aware of all the concomitant drugs used in a particular study, and they may differ from study to study. While the Agency agreed that the intent of such stability assessment is not an all-encompassing and exhaustive investigation, it should be restricted to known or commonly used medications. The Agency indicated that it does not see any reason for not conducting such stability assessments, especially for fixed-dose combinations, but will consider the divergent perspectives.

Multi-Batch Runs

A run can consist of distinct processing batches for various reasons (*e.g.*, several multi-well plates, multiple analysts, limited capacity for extraction). For runs with distinct processing batches, the utility of separate QCs in each distinct processing batch to demonstrate accuracy within those distinct batches, as well as the run, was debated. The Agency described situations wherein the run passed QC acceptance criteria, although the majority of QCs in one or more distinct processing batch(es) within the run failed. Therefore, the accuracy of the distinct processing batches within the run was questionable. This is the basis for the Agency's recommendation to include at least duplicate QCs at all concentrations in each distinct processing batch within a run, and establish acceptance criteria for the whole run, as well as the distinct batches within the runs. While the value of additional QCs in distinct processing batches was recognized, some conference attendees noted that the additional QCs could increase the total number of QCs within an analytical run beyond the generally acceptable 5%. Additionally, it may pose an undue burden when the distinct processing batches have limited sample capacity or when there are insufficient numbers of study samples in each distinct batch. It was also noted that the creation of partial run acceptance criteria had been rejected in previous Workshops due to the complexity of rules and interpretation of the data. There was a discussion on a fit-for-purpose approach; however, no consensus was reached on an approach to assure the accuracy of multi-batch runs.

LIGAND-BINDING ASSAYS: CONSENSUS TOPICS

The group discussed and clarified a number of ligand-binding assay (LBA) issues. The topics for which consensus was reached are described below.

Selectivity/Specificity

The group discussed the particular aspects of selectivity and specificity as they pertain to LBAs; specifically, interferences from substances structurally similar to the analyte, as well as other matrix components should be evaluated. The former is typically related to the specificity of the assay reagents and the latter to the ability of the assay to selectively measure the analyte of interest in a complex biological matrix.

Since small molecule concomitant medications are not structurally similar to large molecule therapeutics, these typically do not need to be tested for interference in the assays. Other substances that may bind the therapeutic molecule (*e.g.*, soluble target) should also be evaluated as appropriate. In the event that a significant interference is identified, development of an orthogonal method may need to be considered. It is, therefore, recommended to test for anticipated interferences during assay development so that method modifications, or development of new methods, may be implemented prior to assay validation and implementation for sample analysis. This would include special populations when known at the time of method development.

The approach for the selectivity assessment to address matrix effects was agreed as previously published (10). Briefly, matrix samples from 10 or more individuals should be tested unspiked and spiked with analyte at LLOQ and HQC concentrations. Eighty percent of samples should meet the following criteria: unspiked samples should measure BQL, and measured concentrations should be within 25% of nominal for LLOQ spikes and 20% of nominal for HQC.

Calibration Curve

The calibration curve should be assessed with calibrators that are prepared independent of the QCs. Method validation experiments should include a minimum of six runs conducted over several days with at least six non-zero calibrator concentrations that span the anticipated quantitative assay range. These concentrations should include LLOQ and ULOQ, but should not coincide with concentrations of the low, mid, or high QCs. Additional calibrators outside the quantitative assay range, referred to as anchor points, may also be included to facilitate curve fitting. Acceptance criteria for the calibration curve during method validation align with previous recommendations (11). Specifically, 75% of the calibrators in the curve should have back-fitted concentrations with a percent deviation ±20% (25% at the LLOQ) (*versus* the stated nominal concentrations) in order to pass acceptance criteria. The number of calibrators should be adequate to accommodate the curve fit model (12). In order to accommodate these criteria, eight calibrators are frequently assessed to ensure that runs with 75% passing calibrators still retain six passing calibrators. It is acknowledged that it may not always be possible to include six to eight calibrators within the assay calibration range.

Run Acceptance

As discussed under "Chromatographic Assay," method accuracy and precision acceptance criteria should be determined from QC performance during validation independent of *a priori* criteria. To accommodate this, run acceptance during A&P validation runs should be based on calibration curve performance only, where calibrator concentrations meet acceptance criteria for (75% of the calibrators in the curve should have back-fitted concentrations with a percent deviation ±20% (25% at the LLOQ) *versus* the stated nominal concentrations to pass). All QCs from passing A&P runs are then evaluated to determine the assay accuracy and precision criteria to be used in subsequent runs and study sample analysis.

Precision and Accuracy

To assess precision and accuracy, five QC concentrations should be used (LLOQ, low, mid, high, and ULOQ) (10,13). Six independent assay runs with at least three replicates of each concentration per run should be conducted over several days. This enables a more robust estimation of both intra- and inter-assay A&P. During validation, the intra-run and inter-run accuracy and precision of the mean concentrations should be within 20% of the nominal value at each level (25% at the LLOQ). These same QCs should also be used to determine the total error. The total error (*i.e.*, sum of absolute value of the % relative error and % imprecision) should not exceed 30% (40% at LLOQ) (10).

Dilutional Linearity and Hook Effect

Dilutional linearity and the hook effect are typically assessed separately from accuracy and precision. Dilutional linearity experiments are performed to demonstrate that high concentrations of the analyte can be accurately measured by diluting into the quantitative range of the assay and multiplying the measured concentration by the dilution factor. The prozone (hook effect) is typically assessed in the same experiment by measuring samples spiked with very high concentrations of analyte without dilution. A prozone is identified when increasing analyte concentration results in no change or decreased signals when compared to the preceding concentration.

Sample Analysis

Calibration curve ranges for LBAs are frequently narrow (1–3 logs) and cannot cover the anticipated range of study sample concentrations for most studies. Instead, sample dilution is employed, supported by demonstration of dilutional linearity during pre-study validation. While there may be circumstances where analyzing all study samples from a single subject in a single run (plate) may be desirable, it is not always possible or practical for ligand-binding PK assays.

Ligand-binding assays generally involve duplicate analysis of all standards, QCs and samples. When newer technologies are not as reproducible compared to a well-characterized LBA, the tendency is to perform triplicate analysis to increase confidence in the results. As LBAs have advanced, particularly in the quality of the assay reagents, it may be justifiable to move from duplicate analysis to singlet analysis as currently implemented for LC-MS/MS methods. Agreement on running samples in singlet from the industry perspective was achieved, as long as QC intra- and inter-plate precision is achieved according to the method validation. Some proposals on tightening these criteria to allow a higher confidence in running singlets were discussed without any concrete conclusion.

LIGAND-BINDING ASSAYS: NON-CONSENSUS TOPICS

ULOQ Acceptance Criteria

Setting ULOQ criteria was an issue that received appreciable discussion. The Agency had proposed an upper limit of 20%, based on review experience. However, counterarguments were made that matching ULOQ to LLOQ criteria enables utilization of the entire curve and removes the need to "artificially" curtail the curve at the upper end to meet tighter criteria. Furthermore, this practice may also reduce the extent of sample dilutions needed, as well as number of potential sample repeats due to out of range results (above limit of quantitation). However, as a counterpoint, it was noted that samples that are above assay range are typically diluted into the midrange and, therefore, results can be read from the most accurate portion of the curve.

Sample Analysis—Placement of Quality Controls

Robust discussion occurred around the proposal to include additional QCs when study sample concentrations are clustered in a narrow range of the standard curve. Opinions varied depending upon where on the curve the sample results clustered. For example, if the majority of sample results clustered between LLOQ and LQC, then there was some agreement that an additional QC to bracket the results might be considered.

It was acknowledged that most LBAs would not require additional QCs, as the QCs are already "close" to one another due to the limited range of the curves. There was also some agreement that the bulk of sample results should be bracketed by two QCs, although for newer technology assays with larger dynamic ranges, there was no agreement on "how close" the bracketing QCs should be.

COMMON ISSUES: CONSENSUS TOPICS

Many of the parameters and principles of bioanalytical method validation are common regardless of assay platforms. This section clarifies recommendations on common topics in bioanalytical method validation and analysis, and highlights consensus on recommended approaches to enhance the quality of bioanalysis for regulatory submission.

Stability

Stability evaluations should cover the expected sample handling (*e.g.*, freeze-thaw cycles), analysis, and storage conditions (*e.g.*, long-term, bench top, and room temperature storage) likely to be encountered during the conduct of the study through analysis, including conditions at the clinical site, during shipment, and at all secondary sites. If actual conditions are known to deviate from expected conditions, or if storage conditions of study samples are changed or exceed validated conditions, additional testing may need to be conducted to verify sample integrity under actual conditions.

Matrix-related stability experiments should be conducted by comparing the experimental stability samples against freshly prepared calibrators and/or QCs. To be considered freshly prepared, calibrators and/or QCs should be prepared (*i.e.*, spiked) on the day of the stability experiment and not frozen prior to use. In addition, it is recommended that the stock solutions used to spike the matrix in preparing stability QC samples is prepared on the day of use or is within its known stability period. While use of freshly prepared calibrators and or QCs is the preferred approach, it is recognized that in some cases, for macromolecules, it may be necessary to freeze them overnight. In such cases, valid justification should be provided and freeze/thaw stability demonstrated.

Matrix-related stability assessments should be conducted with at least three replicates at both low and high concentrations within the validated range of the assay. Assessments of analyte stability should be conducted in the same matrix as that of the study samples, except in cases of rare matrices where suitable surrogate matrices may be used. Matrix-related stability results should be within ±15% of their nominal value for chromatographic assays and within ±20% of nominal for ligand-binding assays. Final study results should not be reported until matrix stability is established.

Long-Term Stability

As previously determined, the storage time in a long-term stability evaluation should equal or exceed the longest time between collection of a sample and the subsequent analysis of that sample.

Stock Solution and Working or Substock Solution Stability

The stability of stock solutions of analytes (drugs, metabolites, and biomarkers) should be evaluated. When the stock solution exists in a different state (solution *vs.* solid) or in a different buffer composition (generally the case for macromolecules) from the certified reference standard, the stability data on this stock solution should be generated to justify the duration of stock solution storage stability. It should be noted that solutions in other solvents and buffers prepared from a stock solution may have different stability characteristics and as such need to have their stability determined.

Processed Sample Stability

Where applicable and according to the technique used, the stability of processed samples, including the resident time prior to analysis (*e.g.*, in an autosampler), should be determined by analysis against freshly prepared calibrators and/or QCs.

Pharmacokinetic Considerations

Concentrations in study samples should not be extrapolated below the LLOQ (lower limit of quantitation) or above the ULOQ (upper limit of quantitation) of the standard curve. Instead, the standard curve should be extended and revalidated, or samples with higher concentrations should be diluted and reanalyzed. Although the draft revision recommended reporting concentration results below the lower limit of quantitation as zero (0), the group generally felt it was more appropriate to report the data as BQL or <LLOQ, which the Agency will consider further.

Repeat analysis of study samples should be conducted in accordance with an established SOP or guideline for the

analysis, acceptance criteria, and selection of final reported values. This SOP or guideline should explain the reasons for repeating sample analysis. Reanalysis conducted for assignable cause could include sample processing errors, equipment failure, or other contemporaneously documented causes that invalidate the original assay(s). Reanalysis conducted for an assignable cause should be conducted with at least the same number of replicates as originally tested. Reanalysis conducted for confirmatory purposes such as apparent measureable drug concentrations in a placebo sample or for inconsistent pharmacokinetics should be based on an established SOP or guideline. Reanalysis of samples for PK reasons in bioequilvalence (BE) studies is discouraged.

Incurred Sample Reanalysis

ISR is now well established as an important element of bioanalysis and is intended to verify the reliability of the reported subject sample analyte concentrations. ISR is conducted by repeating the analysis of a subset of subjects' samples from a given study in a separate run(s) to critically support the performance of assays. To minimize variability, the original and repeat analysis is conducted using the same bioanalytical method procedures. If a bulk frozen calibration curve was used for the original analysis, then it is acceptable to use a frozen curve for the ISR evaluation. The curve, QCs, and study samples for the ISR evaluation are extracted or processed separately from those used in the original runs. For nonclinical safety studies, ISR should be conducted at least once for each method and species.

Regarding selection of samples for ISR, it is not necessary to repeat all time points in the PK profile from selected subjects, but the profile should be adequately represented and include assessments around C_{max} and in the elimination phase for all analytes (*i.e.*, drug and metabolites) for the subjects chosen for ISR evaluation.

As previously defined at the CC IV Workshop on ISR (8), at least two thirds (67%) of the repeated sample results should have a percent difference within 20% for small molecules and 30% for macromolecules. While industry recommends for chromatographic assays of biologics appropriate criteria should be used based on the performance of the method during validation, there was no consensus reached with the Agency. The percent difference of the results is determined with the following equation:

$$\frac{(\text{Repeat}-\text{Original}) \times 100}{\text{Mean}}$$

Standard operating procedures should be established and followed to address sample selection and acceptance criterion for ISR. In addition, written procedures should be established for investigation of ISR failure for the purpose of resolving the lack of reproducibility (*i.e.*, less than 67% of samples achieving performance criteria). All aspects of ISR evaluations, including investigations, should be documented to reconstruct the study conduct. ISR results should be included in the final report of the respective study.

Documentation and Reporting

The workshop report from the 3rd AAPS/FDA Bioanalytical Workshop (9) includes a comprehensive table detailing documentation desirable at the analytical site and in validation and analytical reports. The following items regarding documentation and reporting were clarified and summarize specific discussion points from the current Workshop:

Chain of Custody

Regarding chain of custody of study samples, the documentation of performing laboratories should specify the expected sample collection and storage conditions and as possible, provide to the entities (*i.e.*, the nonclinical and clinical sites and central laboratories) the conditions related to collecting, storing, and shipping samples prior to receipt at the laboratory. Additional testing may be necessary if there are known deviations from the expected conditions.

Accuracy and Precision

When determining accuracy and precision, it is acceptable to exclude QC data that is deemed invalid due to an assigned cause (*e.g.*, documented processing errors, documented equipment failure). However, data for non-assignable causes (*e.g.*, outliers) should not be excluded. If data are significantly impacted by outliers, results excluding the outliers are also recommended to be included in the report for consideration. As stated earlier, during method validation, acceptance of precision and accuracy runs should not be based on QC acceptance, only on the performance of the calibration curves. However, for in-study performance, precision and accuracy estimation is based on QC data from successful runs that meet the acceptance criteria for both calibration curves and QCs.

Reporting Method Development

A robust discussion occurred on whether method development activities should be in scope when reporting results of a validation. It was agreed that a synopsis of method development activities should be included in the validation report, specifically on aspects that are unique (*e.g.*, use of stabilizers). It is not necessary to include a detailed description of all method development data in the validation report. However, when initial validation attempts fail and iterative method modifications are needed or there is significant evolution of the method, these should be captured in the relevant validation report. FDA representatives also indicated that in order to relate data from different methods/studies, submissions (*e.g.*, Section 2.7.1 of the common technical document) should provide a history of the method and where to go for additional information.

Other Items

With regard to preparation of QCs and calibrators, summary information is sufficient to be included in the validation report. Detailed procedures for preparation of QCs and calibrators must be available in the laboratory records.

For chromatographic assays, reintegration of individual chromatograms with different integration parameters than used for the overall run should be avoided. For studies of high importance to a filing (*e.g.*, bioequivalence/biocomparability, pediatric, longest term toxicology studies), if reintegration of some sample within a run is necessary, both the initial and reintegrated chromatograms should be included within the report. In all other studies, if differential reintegration of some samples is performed, it should be noted in the report. In all studies, the original and reintegrated chromatograms should be maintained by the laboratory.

The bioanalytical report should document expected samples that were not received and state the reasons if known. If the samples are "missing" or were mishandled at the bioanalytical facility, such samples should be noted in the report with proper justification.

As stated previously, criteria for reanalysis of samples should be defined in laboratory SOPs. If reanalysis is not per SOP, documentation of justification and manager approval should be maintained within the laboratory records.

Analytical data and tables should be submitted in a clear, easily reviewable format. Inclusion of hyperlinks and/or bookmarks is recommended to simplify location of reported information.

COMMON ISSUES: NON-CONSENSUS TOPICS

Method Validation

Defining specific criteria for the level of validation required to support changes to a previously validated assay was beyond the scope of the Workshop. The extent of validation required during assay transfers and/or changes should use good scientific judgment and should be recorded in SOPs or study validation plans.

Linking Expiration of Reference Standards to Stock Solution

A specific area without consensus between industry and the Agency regarded the draft guidance language around expiration dates of reference material influencing the expiration dates of stock solutions made from the reference standard supply. The industry perspective proposed reference standard stability is not linked to subsequent stock/working solutions; as was discussed and reported in the Crystal City III conference report (9). Supporting rationale for this position included consideration of physical phase changes of the reference standard when going from solid to solution, the logical extension to QC samples spiked with the stock/working solutions and reference to other current regulatory practice including GMP. The Agency agreed to take the industry consensus into consideration based upon the active discussion around the topic.

Incurred Sample Reanalysis

No consensus was reached on the FDA's 2013 draft guidance recommendation that ISRs should be performed on 7% of incurred samples in the study regardless of size. Industry recommends an ISR sample size of at least 5% of the total study samples, and the Agency recognizes that 7% is not consistent with other health authority requests.

A new option was presented for submissions containing only a few studies. If the sponsor chooses, ISR analysis may be incorporated into the method development and validation stage prior to the pivotal study through testing with samples from a pilot study. This approach allows for the remediation of methodological issues prior to conduct of the pivotal study.

The draft guidance indicates that ISR is expected for all *in vivo* human BE studies and all pivotal PK or PD studies. Regarding ISR for PD studies, the Agency indicated that ISR is recommended if the measurement of the PD endpoint requires quantitative bioanalysis. Although further consideration on this topic was slated for later discussion, the topic was not thoroughly discussed.

NEW BIOANALYTICAL AREAS

New Technologies: Consensus Topics

In the draft guidance, the Agency described the need to compare new technologies with older ones that have been used in regulatory submissions, and which they may likely replace or supplement for that purpose. The example described was the use of dried blood samples to replace whole blood or plasma samples during the development of a new pharmaceutical. The purpose is to provide a comparative basis for the data generated by these methods. However, in the broader discussion of this topic, it was noted that some technologies would not be directly comparable for a variety of reasons. For example, with respect to the measurement of macromolecules, some newer technologies are not based on affinity interactions between the analyte and capture/detection reagents, involve a different readout, utilize amplification of signal, or do not reach equilibrium, which makes it difficult to compare absolute analyte concentrations. It was generally agreed that a new method/technology/platform as a stand alone or in combination with automation would be acceptable for use in regulated studies and would need to comply with the expectations for method validation and sample analysis.

However, different methods, technologies, and platforms may not provide the same absolute analyte concentration. The consensus perspective was that bridging data between methods would be the more correct way to proceed when changing methods for large molecules encompassing a set of spiked samples measured with the two different methods. The differences expected and/ or observed between the two methods should be explained with the appropriate scientific rationale and the expectations for interpretation and use of the data from the two assays.

With the onset of new technologies, different run operations may also be observed where instead of running the standards and samples in a plate format, they may be run in a series that is not defined by the solid structure (*e.g.*, microtiter plate) used for running the standards/samples. No consensus was obtained with the size of the batch run and this topic is currently left open. On the other hand, it was agreed that the analytical run was defined as consisting of a set of standards with interspersed QCs on each solid support used for running the standards/samples. As previously required, the FDA recommends using a number of QCs that are equal to or greater than 5% of the total unknown samples in a run. This can be useful for runs that are not defined by the solid support and for alternative formats such as 384-well plates.

Other minor aspects of some of the instruments used in combination with new technology may involve automated or semi-automated systems and consideration should be given to the type of methodology applied in developing appropriate validation testing. Also, carryover may be seen with some technologies with reusable tips or liquid-flow paths, and carryover should be assessed and avoided in the development phase. If it cannot be avoided, a rationale on how to deal with carryover should be recorded in the validation report and in place before performing sample analysis.

Antibody Drug Conjugates

ADCs combine a therapeutic small molecule drug with a cell-targeting specific antibody (carrier), and it is anticipated that antibody-based specific delivery of the drug molecule to targeted cells should improve efficacy and safety.

The measurement of several different analytes has been proposed to fully evaluate the correlation between ADC therapeutic exposure and the observed efficacy and safety signals (14). Both LBA and LC-MS/MS platforms have been broadly utilized. It has been demonstrated that the assays used in support of ADCs, *e.g.*, LBA protocols designed to measure conjugated and total antibody analytes, are often sensitive to the drug antibody ratio (DAR value) (15,16). Furthermore, as payload molecules are released during systemic circulation, the composition of ADC species in study samples will generally not be equivalent to that found in the original reference material. Ideally, the assay should be developed to accurately detect appropriate ADC species found in circulation.

LC-MS/MS-based protocols have been applied to evaluate concentrations of unconjugated and antibody-conjugated drug (17). Concentrations of unconjugated drug are generally expected to be very low, requiring highly sensitive assays. These assays should be highly sensitive, and capable of stabilizing the ADC and preventing post-collection deconjugation. Typical performance verification tests and criteria applied in the validation of assays used in support of biotherapeutic compounds are expected to be relevant for ADC-related bioanalytical methods (3,9,10). This includes expectations regarding assay precision and accuracy, specificity, selectivity, and robustness. Regular A&P acceptance criteria should apply with some significant considerations presented below.

As ADCs are still a relatively new modality, and there is limited regulatory experience, the Agency declined to concur with the industry position on these molecules at this time. As a general approach, the industry is proposing assay acceptance criteria based on the assay performance data obtained during method validation for LC-MS assays or LBAs. Similarly, the acceptable accuracy range for complex biotherapeutic modalities and/or LBA methods requiring complex sample preparation or procedural steps may need to be adjusted from the current generally applied criteria with a consideration of how the data will be used. Final assay acceptance criteria for LBA protocols should be based on the assay performance data obtained during method validation.

Stability tests are expected to mimic typical handling conditions including sample handling, freeze/thaw, bench top, and long-term storage stabilities. Stability of active catabolites released from ADC (*i.e.*, unconjugated drug or linker-drug) should be tested in a spiked matrix as is typically done for conventional therapeutics. Stability of ADC catabolites may need to be evaluated in the presence of the dosed ADC to assess potential release of catabolites during post-collection storage and sample handling. Because unconjugated drug represents only a small fraction of the total (conjugated and unconjugated) drug found in the sample, even a minor instability of the intact ADC during sample collection and manipulation may lead to a substantial artifactual increase in measured unconjugated drug concentration, thereby impacting the overall accuracy of the test. The industry position is that the acceptable range of variability for analyte stability should be scientifically justified, based on the assay validation data and with a consideration of how the data will be used.

Intrinsic and *in vivo*-induced heterogeneity of ADCs calls for a careful review of the definition of the term reference standard when applied to complex biotherapeutic modalities. Reference material should be a well-characterized, traceable batch of material of known purity and concentration, accompanied by a CoA or similar documentation. Ideally, the reference standard should be as similar as possible to the measured analyte (*e.g.*, with respect to drug antibody ratio distribution and unconjugated drug content). ADC reference material should be used to prepare standard solutions and QCs for total antibody, conjugated antibody, and antibody-conjugated drug analytical tests.

BIOMARKER ASSAYS AND DIAGNOSTIC KITS

Guiding Principles

The Workshop was the first opportunity for a wide discussion of these new areas of consideration in the guidance. Biomarkers entail a very broad range of entities, and the discussion centered on fundamental issues with the recognition that this was the beginning of the dialogue. Biomarkers are measured by a broad group of highly sophisticated technologies that include immunoassay, mass spectrometry, flow cytometry, molecular pathology, and genomics to name a few examples. While the recommendations and general fit-for-purpose concepts contained herein are applicable to all biomarker technologies, the FDA draft BMV guidance and subsequent discussions held at the CCV meeting focused only on immunoassay and mass spectrometry. Likewise, biomarkers encompass a broad range of measurements that include pharmacodynamics, target engagement, receptor occupancy, and safety to name a few. This discussion will be limited to non-safety biomarkers, as safety biomarkers were not extensively discussed at the Workshop. Lastly, as stated in the draft BMV guidelines, the recommendations pertain mainly to the validation of assays to measure *in vivo* biomarker concentrations from common biological matrices such as blood or urine and their use in study sample analysis.

Biomarker Categories

In developing biomarker assays, it is useful to think of these assays as belonging to one of two broad categories based on the intended use of the data. Category 1 contains those assays that generate data for internal decision making and do not drive label claims. In contrast, category 2 contains those assays that generate data to be used to support pivotal determinations of effectiveness or label dosing instructions and therefore, it is critical to ensure the integrity of the data. These categories are not always distinct

entities and during drug development, a spectrum of validation will typically occur.

Category 1 assays tend to be exploratory endpoints where the data can help the sponsor understand the pharmacodynamics or mechanism of action of the compound or more broadly, for hypothesis generation. These assays can vary greatly in quality and robustness from non-FDA-approved assays (in-house developed assays, commercial-research use only kits, and lab-developed tests) to FDA-approved diagnostic (Dx) kits run in a central clinical lab. These assays can be definitive, relative quantitative, or qualitative.

It is important to establish a consistent language around these assays, and the following definitions were recommended and are presented in order of increasing levels of validation characterization required (18).

- Qualitative Assay: The assay readout does not have a continuous proportionality relationship to the amount of analyte in a sample; the data is categorical in nature. Data may be nominal (positive or negative) such as presence or absence of a protein, or ordinal, with discrete scoring scales (1 to 5, −/+, +++, *etc.*).
- Relative Quantitative Assay: A method which uses calibrators with a response-concentration calibration function to calculate the values for unknown samples. The quantification is considered relative because the reference standard is either not well characterized, not available in a pure form, or is not fully representative of the endogenous biomarker.
- Definitive Quantitative Assay: An assay with well-characterized reference standards, which represents the endogenous biomarker, and uses a response-concentration calibration function to calculate the absolute quantitative values for unknown samples.

Category 2 assays must be able to generate data of high integrity and therefore, one must take into account the different quality of assays available (non-FDA-approved and FDA-approved) and validate appropriately. It is recognized that obtaining appropriate biomarker reference standard or blank matrix may not always be possible. Additionally, a reference standard may not be identical to the endogenous biomarker. As a consequence, developing and validating the biomarker assay cannot always be performed to meet the standards of a PK assay, although it is recommended to meet the criteria as closely as possible. In these cases, the approach described by Lee et al. (18) may be useful in achieving these goals.

For a category 2 assay, if a definitive quantitative assay can be developed, then the assay should be validated to meet the same standards of a PK assay. In contrast, if only a relative quantitative assay can be developed due to technical limitations of the matrix and/or reference standard, then the assay should be validated and assay acceptance criteria should be set with a view to approaching as closely as possible the expectations for a PK assay. Industry recommended the preparation of QC samples; specifically using QC ranges prepared by pooling low/medium/high biomarker containing matrix and subsequently used as analytical QCs for assay run acceptance. This is an area requiring further discussion.

For all category 2 assays, every effort should be made to evaluate critical assay parameters (*e.g.*, standard curve performance, accuracy and precision, sensitivity, stability, specificity, parallelism) to make the assay as close to a PK assay as technically feasible. Where an assay parameter does not meet the criteria of a PK assay, it is recommended that the sponsor provide an explanation for why this should not pose any issues in data quality for its intended use. Furthermore, early interaction with the Agency to discuss these issues may be helpful.

Multi-analyte analysis is commonly practiced with biomarkers. In many cases, the method is validated for all analytes simultaneously, but in some cases, LBAs for multi-analytes are validated individually and then combined together in the method validation. Once in sample analysis, if one of the analytes does not pass their criteria, the sample should be rerun under the same conditions and the other previously passed analytes not taken into consideration. The FDA did not oppose this option.

Use of FDA-Approved Dx Assays

Special consideration must be taken into account when using an FDA-approved Dx kit to generate results with a category 2 assay. The main challenge with Dx kits is the assays are often on closed platforms and cannot be further optimized (*i.e.*, the sponsor cannot add additional standard points or change assay buffer). Therefore, Dx assays with sparse calibration standards should include additional validation experiments using additional standards to evaluate the calibration range and standard curve performance. Similarly, other aspects of the validation (specificity, selectivity, stability, *etc.*) may require supplemental work. The goal and benefit of conducting these additional experiments is to generate a sufficient body of evidence that support using the Dx kit, as intended by the Dx manufacturer, will deliver the data quality required during in-study sample analysis.

While these Dx assays are usually highly validated and follow College of American Pathologists (CAP) Clinical Laboratory Improvement Amendments of 1988 (CLIA) guidelines, the assay simply may not be fit-for-the-purpose in clinical development. These assays were specifically developed and validated for the purpose of diagnosing a patient health/disease status. Therefore, some parameters that are critical to how samples are evaluated in the context of a clinical study may not be sufficiently evaluated as part of the Dx assay validation package. Here, it is the responsibility of the sponsor to perform additional verification experiments to augment the validation package and demonstrate that the assay is suitable for the intended use. For example, the laboratory could conduct the appropriate stability experiments that provide data to support how study samples were collected and stored. One strategy to evaluate stability is to prepare and analyze mock clinical samples that mimic the way samples are collected, stored, shipped, and analyzed. The concern here is that the Dx assay may have been validated to analyze samples that have been stored, for example, at ambient temperature for less than 48 h. If the clinical samples will be collected, stored, shipped, and analyzed under these conditions, then there are no issues, and the assay can be used as is. However, if the samples are stored frozen and/or analyzed after 48 h, then it is up to the

sponsor to determine if the assay is fit for purpose. The same would hold true for the evaluation of potentially interfering compounds to ensure the assay still meets specificity requirements or differences in anticoagulant.

Biomarkers represent a growing and important data set to support and drive approvals of new drugs. As such, the above discussion establishes an initial position for industry and the agency to work from, recognizing that the rapidity of change will require an ongoing dialogue and flexibility as new technology is applied and other approaches mature.

CONCLUSION

The Crystal City V Conference was a forum for interaction between the pharmaceutical industry and regulatory agencies. Extensive discussions were held for a wide range of bioanalytical topics for chromatographic and ligand-binding assays. Consensus was reached on a variety of subjects for these platforms, which was also consistent with the majority of decisions from CCIII and CCIV workshops. Several new topics in bioanalysis were also discussed, namely biomarkers, diagnostic kits, ADCs, and mass spectrometry of large molecules. Biomarkers and diagnostic kits, which are the latest additions to the FDA BMV Guidance, were robustly discussed, and the challenges and limitations of the assays were presented. Bioanalytical assays and validation specifications, needed for antibody-drug conjugates, were also presented and discussed. Quantification of large molecules by mass spectrometry continues to increase and the bioanalytical challenges are being addressed within the industry. At the current Workshop, this area of science was not discussed due to the lack of broad experience available for a robust discussion. There were some areas of non-consensus, which was usually the result of a lack of general experience with the issue. These issues will serve as the basis for future industry-agency interactions and improvements to the science.

ACKNOWLEDGMENTS

We would like to acknowledge all of the panelists, moderators, and note takers for their contributions to the success of the workshop. Stanley Au, U.S. Food and Drug Administration, Silver Spring, MD USA; Surendra K. Bansal, Roche TCRC Inc., New York, NY, USA; Chris Beaver, inVentiv Health Clinical, Montréal, QC, Canada; Ronald R. Bowsher, B2S Consulting, Indianapolis, IN USA; Margarete Brudny-Kloeppel, Bayer Pharma AG, Berlin, Germany; Christopher Evans, GlaxoSmithKline, King of Prussia, PA USA; Douglas M. Fast, Covance Laboratories, Madison, WI, USA; Chad A. Ray, Pfizer Inc., San Diego, CA; Scott Fountain, Pfizer Inc., San Diego, CA USA; Fabio Garofolo, Algorithme Pharma, Laval, QC Canada; Russell P Grant, Laboratory Corporation of America, Burlington, NC, USA; Michael Hayes, Novartis, East Hanover, NJ USA; Roger N. Hayes, MPI Research, Mattawan, MI USA; Olutosin Remi Idowu, U.S. Food and Drug Administration, Silver Spring, MD USA; Rand Jenkins, PPD, Richmond, VA USA; Marian Kelley, MKelley Consulting LLC, West Chester, Pa, USA; Lindsay E. King,, Pfizer Inc., Groton, CT USA; Johanna Mora, Bristol-Myers Squibb Co, Princeton, NJ USA; William Nowatzke, Radix BioSolutions, Georgetown, TX, USA; Mark Rose, CHDI Management, Inc., Los Angeles, CA USA; Nilufer Tampal, U.S. Food and Drug Administration, Silver Spring, MD USA; Theingi Thway, Amgen Inc, Thousand Oaks, CA USA; Peter van Amsterdam, Abbott, Weesp The Netherlands; Faye Vazvaei, Roche TCRC Inc., New York, NY, USA; Leah Williamson U.S. Food and Drug Administration, Silver Spring, MD USA; Chongwoo Yu, U.S. Food and Drug Administration, Silver Spring, MD USA; Paul Wielowieyski, Health Canada, Canada; Olivier Leblaye, l'Agence Française de Sécurité Sanitaire des Produits de Santé AFSSAPS, France; Noriko Katori, National Institutes of Health Sciences, Japan; Joao Tavares Neto, ANVISA, Brazil.

The views expressed in this article are those of the authors and do not reflect official policy of their individual organizations. No official endorsement is intended or should be inferred.

A special thank you to Elizabeth Scuderi and the AAPS staff who handled all of the background activities to enable a smoothly functioning workshop.

REFERENCES

1. Shah VP, Midha KK, Dighe S, McGilveray IJ, Skelly JP, Yacobi A, *et al*. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Pharm Res. 1992;9(4):588–92.
2. Conduct and analysis of bioavailability and bioequivalence studies—Part A: Oral dosage formulations. [book auth.] Health Canada. s.l. : Ministry of Health, Heath Products and Food Branch. Canada. 1992.
3. Food and Drug Administration, USA. Guidance for industry bioanalytical method validation. [Online] May 2001. http://www.fda.gov/cder/guidance/index.htm.
4. Guide for validation of analytical and bioanalytical methods, Resolution - RE n. 899, of May 29, 2003, Agência Nacional de Vigilância Sanitária www.anvisa.gov.br.
5. EMA, European Medicines Agency. Guideline on bioanalytical method validation. [Online] July 21, 2011. EMEA/CHMP/EWP/192217/2009.
6. Guideline on bioanalytical method validation in pharmaceutical development. Japan: Pharmaceutical Manufacturers Association; 2013.
7. FDA, US Department of Health and Human Services. Draft guidance for industry: bioanalytical method validation (Revised). [Online] September 2013. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM368107.pdf.
8. Fast DM, Kelley M, Viswanathan CT, O'Shaughnessy J, King SP. Workshop report and follow-up—AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples—implications of Crystal City recommendations. AAPS J. 2009. doi:10.1208/s12248-009-9100-9.
9. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, *et al*. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. AAPS J. 2007;9(1):E30–42.
10. De Silva B, Smith W, *et al*. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. Pharm Res. 2003;20(11):1885–900.
11. Viswanathan CT, Bansal S, Booth B, De Stefano AJ, Rose MJ, Sailstad J, *et al*. Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. AAPS J. 2007;9(1):E30–42.
12. Findlay JW, Dillard RF. Appropriate calibration curve fitting in ligand binding assays. AAPS J. 2007;9(2):E260–7.
13. Kelley M, DeSilva B. Key elements of bioanalytical method validation for macromolecules. AAPS J. 2007;9:E156–63.
14. Gorovits B, Alley SC, Bilic S, Booth B, Kaur S, Oldfield P, *et al*. Bioanalysis of antibody-drug conjugates: American Association

of Pharmaceutical Scientists Antibody-Drug Conjugate Working Group Position Paper. Bioanalysis. 2013;5:997–2013.
15. Stephan JP, Chan P, *et al.* Anti-CD22-MCC-DM1 and MC-MMAF conjugates: impact of assay format on pharmacokinetic parameters determination. Bioconjug Chem. 2008;19(8):1673–83.
16. Stephan JP, Kozak KR, *et al.* Challenges in developing bioanalytical assays for characterization of antibody-drug conjugates. Bioanalysis. 2011;3(6):677–700.
17. Alley SC, Benjamin DR, *et al.* Contribution of linker stability to the activities of anticancer immunoconjugates. Bioconjug Chem. 2008;19(3):759–65.
18. Lee J, Viswanath D, Barrett Y, Weiner R, Allinson J, Fountain S, *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. Pharm Res. 2006;23:312–28.
19. Wang L, Amphlett G, *et al.* Structural characterization of the maytansinoid-monoclonal antibody immunoconjugate, huN901-DM1, by mass spectrometry. Protein Sci. 2005;14(9):2436–46.
20. Junutula JR, Raab H, *et al.* Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. Nat Biotechnol. 2008;26(8):925–32.
21. Xie H, Audette C, *et al.* Pharmacokinetics and biodistribution of the antitumor immunoconjugate, cantuzumab mertansine (huC242-DM1), and its two components in mice. J Pharmacol Exp Ther. 2004;308(3):1073–108.
22. Sanderson RJ, Hering MA, *et al.* In vivo drug-linker stability of an anti-CD30 dipeptide-linked auristatin immunoconjugate. Clin Cancer Res. 2005;11(2 Pt 1):843–52.
23. Xu K, Liu L, *et al.* Characterization of intact antibody-drug conjugates from plasma/serum *in vivo* by affinity capture capillary liquid chromatography-mass spectrometry. Anal Biochem. 2011;412(1):56–66.
24. Health Canada. Conduct and analysis of bioavailability and bioequivalence studies—Part A: oral dosage formulations. Ministry of Health, Heath Products and Food Branch. s.l. : Canada. 1992.