AAPS and US FDA Crystal City VI workshop on bioanalytical method validation for biomarkers

Steve Lowes*,1
& Bradley L Ackermann²
¹Bioanalytical Consultant, 418 Ferguson Road, Freeville, NY 13068, USA
²Investigative Toxicology, Eli Lilly & Company, Lilly Corporate Center, DC 0720, IN 46285, USA
*Author for correspondence: stevelowes@ymail.com

Crystal City VI Workshop on Bioanalytical Method Validation of Biomarkers, Renaissance Baltimore Harborplace Hotel, Baltimore, MD, USA, 28–29 September 2015

The Crystal City VI workshop was organized by the American Association of Pharmaceutical Scientists in association with the US FDA to continue discussion on the bioanalysis of biomarkers. An outcome of the Crystal City V workshop, convened following release of the draft FDA Guidance for Industry on Bioanalytical Methods Validation in 2013 was the need to have further discussion on biomarker methods. Biomarkers ultimately became the sole focal point for Crystal City VI, a meeting attended by approximately 200 people and composed of industry scientists and regulators from around the world. The meeting format included several panel discussions to maximize the opportunity for dialogue among participants. Following an initial session on the general topic of biomarker assays and intended use, more focused sessions were held on chromatographic (LC–MS) and ligand-binding assays. In addition to participation by the drug development community, significant representation was present from clinical testing laboratories. The experience of this latter group, collectively identified as practitioners of CLIA (Clinical Laboratory Improvement Amendments), helped shape the discussion and takeaways from the meeting. While the need to operate within the framework of the current BMV guidance was clearly acknowledged, a general understanding that biomarker methods validation cannot be adequately depicted by current PK-centric guidelines emerged as a consensus from the meeting. This report is not intended to constitute the official proceedings from Crystal City VI, which is expected to be published in early 2016.

The Crystal City workshops on bioanalytical method validation (BMV) have provided an ongoing forum for discussion of US FDA guidance and perspective documents by members of the bioanalysis community. Initiated in 1990, these 2–3 day meetings held in the Washington DC area are attended by drug development and regulatory scientists. In contrast to previous Crystal City workshops, Crystal City VI (CCVI) was entirely devoted to a specific topic, biomarker analysis. Over the past few years, practitioners have attempted to apply existing BMV guidelines to biomarkers with various degrees of success. This is understandable in that BMV guidance from the FDA and all other global health authorities have previously focused only on PK evaluations of drug therapies. As explained by FDA representatives, drug filing applications are now routinely including quantitative biomarker data. Under scrutiny, the FDA has seen inconsistencies with the associated bioanalytical validations performed. This led to a decision to include biomarker assays in the latest draft FDA BMV Guidance [1]. The topic was not resolved at

FUTURE SCIENCE fsg part of

CCV [2] and in turn this established the need for the CCVI workshop.

Bioanalysis of endogenous molecules presents a number of inherent challenges including identification of appropriate reference standards, ensuring adequate specificity, and the need for alternative calibration since the target analytes of interest are present in the biological control matrix. Indeed, the FDA has received sufficient feedback from the bioanalytical community on how to address these core issues within the context of the draft FDA BMV Guidance. While FDA cited cases at the meeting of insufficient validation of biomarker methods used to support label claims, they reiterated their intent to approach this subject in a fit-for-purpose manner. To be clear, the FDA is not proposing that all biomarker assays be consistent with fully validated PK methods used to support pivotal trials, but rather that a flexible framework be established that embraces relevant science and helps bioanalysts and regulators alike follow a defendable approach.

The unique nature of biomarker assays and where they fit into regulated bioanalysis set the platform for an engaging discussion at the workshop. The organizers arranged the meeting to encourage discussion by using a format that proved successful at Crystal City V. Four half-day sessions were established covering: general aspects of biomarker assays, chromatographic assays, ligand-binding assays (LBAs) and conclusions of the meeting.

General biomarker overview

The organizers of the workshop opened the meeting with an appeal for active discussion. It was further stated that the objective for the meeting was to advance the topic of biomarker bioanalysis beyond the discussion initiated at Crystal City V. Moreover, the intended purpose of the conference was not to identify specific regulatory language. With the presence of the clinical laboratory scientists in mind, the organizers were quick to establish that the scope of discussion was limited to drug development. While clinical laboratory guidelines and standards were acknowledged as a potentially useful reference point, the need to differentiate routine patient care from drug development was repeatedly stated. Accordingly, clinical diagnostics and laboratory developed tests aimed at patient care should not be confused with drug development and drug approval needs. It was further clarified that drug development includes all activities through phase IV.

After the opening comments, keynote presentations from industry representatives and regulators reviewed the current status of biomarker assays and the challenges presented. The FDA also provided insights regarding their current experience in reviewing data submissions. The talks were followed by an open discussion involving a panel of experienced bioanalysts and regulators who fielded extensive questions from the audience.

The topic of accuracy was introduced early in the discussion. A major factor limiting accuracy is the lack of suitable materials to depict the relevant forms of circulating biomarkers, particularly for proteins. This issue is compounded in drug development owing to the lack of consensus reference materials that are more frequently available in clinical practice. Because of inherent difficulty in achieving truly accurate measurements for endogenous molecules, it was understood that most quantitative biomarker measurements will be made using relative quantitation and not by definitive methods making reference to the definitions from the Biomarker Fit-for-Purpose White paper by Lee *et al.* [3]. With that said, a consensus view from the meeting was that biomarker validation extends beyond measuring precision and that relative accuracy (bias) should be measured whenever possible using the best available standard. While a subset of small molecule LC–MS methods may be considered definitive, it was acknowledged that achieving definitive methods for protein biomarkers is currently exceedingly difficult.

On the topic of fit-for-purpose, early reference was made to the two categories of biomarker assays defined in the Crystal City V conference report [2]. That is, when a biomarker assay is intended to support early drug development, the sponsor should incorporate the extent of method validation deemed appropriate. In contrast, when biomarker assays are used for regulatory action such as pivotal determinations of safety, efficacy or dosage labeling there is an expectation of full validation. These were subsequently defined as Category 1 and Category 2 assays, respectively. This two-tier construct was frequently referenced throughout the meeting with respect to fit-for-purpose and remained relevant and appropriate in the conclusions of the workshop.

The declaration that biomarkers are not drugs was reiterated by a number of speakers in the opening session and throughout the meeting. Outside of endogenous compounds developed as drug therapies, this statement was generally agreed upon. By extension, it was stated several times that biomarker assays are not PK assays. Some participants took this further to suggest biomarker assays need not, and should not, be validated the same way as PK assays. While the premise was largely agreed, it was also acknowledged that PK assay validation has many features that can be directly applied to biomarker assays. The concept of using PK assays as a starting point for biomarker assay validation was revisited throughout the meeting as it pro-

vides important connectivity to the FDA draft BMV guidance [1].

When considering a fit-for-purpose approach to biomarker assay validation, five key questions were presented by the FDA for consideration: what is the purpose of the study?; is what is intended to be measured being measured?; what are the limits of the measurement, that is, what is the biomarker change expected?; how much variability/error is in the measurement?; and how do handling conditions affect the measurement?

Implicit in these questions is the need to understand the underlying biology of the biomarker in question. This knowledge is critical to determining the relevance and use of biomarker candidates and the quantitative requirements of an assay. For example, while the sensitivity required to measure basal levels is typically straightforward, the assay range needed depends on both the magnitude and direction of the anticipated change. This discussion introduced the total allowable error (TAE) concept [4] as an alternative to the fixed and prescriptive assay performance criteria characteristic of PK assays. Considering the varying degree of biomarker concentration changes encountered, it was debated whether the fixed and prescriptive performance criteria (e.g., 4-6-X) are appropriate for such assays. Using the TAE approach, the limits for precision and bias are determined from an analysis of inter and intra-subject variation for the biomarker of interest in the intended population for analysis [5]. An interesting twist on this concept is found in cases where the budgeted TAE is high; due to dramatic fold changes (e.g., >2-fold). In such instances, general reluctance was expressed by the attendees for using exaggerated acceptance criteria when a more accurate and precise assay is readily achievable. Clearly more discussion on the TAE approach and its applicability for setting validation acceptance criteria for biomarker methods is needed.

Chromatographic assays

The second session of the workshop was dedicated to chromatographic assays, predominantly LC–MS techniques. Presentations from industry covering biomarker bioanalytical strategy and practical recommendations were followed by open-floor questions to the industry/regulator panel. There were indications from the discussions that the LC–MS community is open to a standardized approach to biomarker BMV albeit different from PK assay bioanalysis. In doing so the Clinical Laboratory Standards Institute (CLSI) was introduced to the discussion. Existing CLSI guidelines for diagnostic marker assays were referenced as pertinent documents that could help formu-

late an approach to drug development needs. It was not proposed that the industry adopt the CLSI guidelines but rather use them as a starting point to develop practices that are appropriate for drug development.

Consistent with the preceding session, extensive discussion continued on the need to measure relative accuracy (bias) in cases where the reference material does not ideally match the endogenous biomarker. Because protein biomarkers often lack a good quality reference standard, the concept of using statistically characterized healthy and diseased state pools to support quantitation was introduced. Proposed was the use of the disease pool as the ULOQ and the healthy pool as the LLOQ (in cases where the biomarker increases with disease). By combining the endogenous pools in ratios of 3:1, 1:1 and 1:3 this admixture approach produces a five-point calibration curve as previously published [5]. Further dilution of the healthy pool with analyte-free surrogate matrix extends the calibration and redefines the LLOQ accordingly. Admixing is dependent on availability of sufficient individual lots of healthy and disease state matrix that can be subsequently pooled to support the resulting relative accuracy assessments. As such, no consensus was reached on the practicality of this approach in the drug development arena, but it remained an attractive consideration for further review.

Contrasting the admixing approach came discussion around spiking reference standard to a surrogate matrix or use of a surrogate analyte [6]. Both approaches require a parallelism experiment that was considered essential to method validation. That is, demonstration that the surrogate used for quantitative calibration appropriately correlates with the endogenous matrix or analyte. When a well-characterized reference standard is available, as with many small molecule biomarkers, using spiked calibrators offers definitive quantitation as long as assay parallelism can be demonstrated. The typical linearity of LC–MS supports parallelism assessments and is also important to demonstrating dilutional linearity below the endogenous level.

Analytical QC preparation should correspond with either the admixing or the spiked reference standard approach taken for the calibrators. As per PK assays, assay performance is determined from the QCs, however, there was general agreement that precision and accuracy assessments should not follow a fixed and prescriptive 4-6-X acceptance criteria rule. There was consensus that endogenous pool QC samples should be used throughout prestudy and in-study method validation. Endogenous QCs allow important inter- and intra-assay assessment during validation of the bioanalytical method and study sample analysis.

The discussion around method validation stability experiments also emphasized endogenous QCs including healthy and disease state pools. Inherent to the endogenous analyte is the inability to establish a true 'time = 0' (T_0) in effect again requiring a relative bias approach. Spiked QCs do support a T_0 comparison if allowing for freshness of the preparation as has become precedent with PK bioanalysis. The stressed conditions of sample storage, freeze-thaw and in-process stability were agreed as relevant to biomarker assay validation. However, assessment of stability of the biomarker in the preanalytical phases, that is, sample collection, handling and shipping stability did not reach consensus. It is difficult for the bioanalytical laboratory to manage this issue beyond providing advice derived from the development and validation of a bioanalytical method. Of course, biomarkers are susceptible to changes during the preanalytical phases and without a handle on such stability there will be influence on assay accuracy. Considering the regulatory interest in preanalytical events and their impact on reported biomarker concentrations, it is expected that this topic will need further discussion and resolution.

The challenges associated with comparing bioanalytical data sets was discussed from multiple perspectives during the workshop. Continuity of the reference standard used when it may be relatively uncharacterized and can potentially vary lot-to-lot is one factor to consider. Likewise, differences in immune-capture reagents or endogenous matrix can result in interassay variability. To bridge assays and studies over time, the use of pooled endogenous and QC samples across methods was recommended. Referred to as 'longitudinal pools' such controls are considered important to data comparisons with emphasis on achieving assay consistency as opposed to accuracy. This applies across reference standards, reagents, control matrices and the laboratories conducting the assays. However, the appeal for consistency in assay performance did not extend to general support of incurred sample reanalysis (ISR). Preference for endogenous, disease state pools analyzed in each batch over ISR was noted in the workshop conclusions.

LBAs

Similar to the preceding discussion on chromatographic assay topics, the third session contained industry presentations followed by an open discussion with a panel of experts. Some of the challenges and concerns were by now familiar including obtaining well-characterized reference standards, defining assay accuracy, establishing assay range and bridging between datasets. Because of the historic position of LBAs as the default method for protein biomarker analysis,

issues brought up in previous sessions related to protein reference standards and methods for QC preparation featured prominently in this session. A further emphasis on differentiating biomarker assays from PK assays emanated from the LBA representatives who also stressed the need to understand the biology at all stages when making decisions about the purpose and use of biomarker data.

In designing LBA method validation experiments, the notion of 'learn and confirm' was used to describe the relationship between method development and validation. While experiments such as specificity assessment typically occur during method development, establishing parallelism begins during method development and becomes an essential feature examined during validation. In the context of category 2 assays, the following attributes are recommended for validation: precision, sensitivity, range, parallelism and endogenous analyte stability. It is important to note that relative accuracy was also included in this list.

Throughout the LBA session reference was made to the importance of understanding the precision of the assay and its relationship to total allowable error. Ultimately, the sensitivity and precision required depend on the biology and the change in biomarker concentration that needs to be measured. It was further recommended that assessments of healthy and disease state samples be taken into account. Typically, these matrices can be commercially acquired or, in some cases, may be obtained from predose samples acquired in clinical trials. While specific assay acceptance criteria were not defined, it was stated that it is a basic expectation that the precision achieved in biomarker assays will meet or exceed current PK-related acceptance criteria.

Assay parallelism was also designated as key to LBAs. For immunoassays, parallelism occurs when the biomarker analyte and the reference standard calibrator are immunologically similar, meaning that the assay detects them the same, regardless of differences in matrix composition. Dilution of the sample with appropriate surrogate matrix to achieve assay parallelism, while still meeting sensitivity requirements, is standard LBA practice [7]. Defining the minimum required dilution (MRD), established in method development and demonstrated in method validation, is thus critical to LBA assays. With respect to assay sensitivity, several options for LLOQ assignment were mentioned, although a consensus position was not articulated.

As with chromatographic assays, the use of endogenous analyte was recommended to define assay performance in validation and to manage changes during use of the assay. The lot-to-lot changes of LBA reagents were frequently referenced as being problematic and a mes-

sage of qualifying reagents and bridging across lots featured strongly in the discussion session. Longitudinal QCs were recommended to assist in this process.

The topic of stability also had familiar themes including the importance of having an endogenous QC sample and the high priority given to understanding and controlling for preanalytical variables. It was acknowledged that for LBA analyte instability is only detected through a change in epitope binding. Therefore, the results may differ according to the reagents used and may also differ from LC–MS results. Lastly, on the subject of ISR, the LBA community shared the consensus of the LC–MS group in that ISR should not be required for biomarker analysis. A preference for using well designed longitudinal controls to understand assay repeatability was voiced.

Conclusion

The Crystal City VI meeting was convened to continue a dialogue initiated at Crystal City V about bioanalytical practices for biomarkers. The format of the meeting, coupled with the cross-section of participants, led to robust discussion on several important topics. Because of inherent difficulty in achieving truly accurate measurements for endogenous molecules, it was quickly understood that most quantitative biomarker measurements will be made using relative quantitation and not by definitive methods [3]. Another consensus view from the meeting was that biomarker validation extends beyond measuring precision and that accuracy or relative accuracy (bias) should be measured whenever possible using the best available standard.

It was also understood that challenges associated with biomarkers, such as difficulty in obtaining viable reference standards and the need to demonstrate parallelism, are magnified when analyzing proteins over small molecules. While the FDA has requested standard practice through their draft BMV Guidance [1], it was evident from the CCVI workshop that guidelines developed for PK cannot be directly applied to biomarkers. It was also clear that the bioanalytical community welcomes scientific discourse and flexibility when discussing how specific biomarker assays should be validated. While all agreed on the need to work within the current regulatory framework provided by the draft guidance, further discussion is needed to create a mechanism that embraces scientific flexibility, yet establishes sufficient prescriptive guidance for effective regulation.

Input from the clinical laboratory scientists provided a consistent point of reference during discussions held at the conference. As the concluding comments included 'let's not re-invent the wheel,' it can be expected that further review of CLSI guidelines is warranted. At the same time, the unique needs of drug development must be featured prominently in this ongoing discussion.

The CCVI workshop achieved the objective of furthering the discussion that was necessary following the previous Crystal City meeting and the FDA draft Guidance language. The forum allowed open discussion between the stake-holders present and highlighted areas that are common to all as well as a better appreciation of the challenges unique to each. Consensus topics were noted while areas that need continued development were identified for follow-up actions. These are expected to be documented in the official proceedings of the workshop which will be published soon.

References

1 FDA Draft Guidance for Industry, Bioanalytical Method Validation.
www.fda.gov/downloads/drugs/
guidancecomplianceregulatoryinformation/guidances/
ucm368107.pdf

2 Booth B, Arnold M, DeSilva B *et al.* Workshop Report: Crystal City V – quantitative bioanalytical method validation and implementation: the 2013 revised FDA guidance. *AAPS J.* 17(2), 277–288 (2014)

3 Lee J, Devanarayan V, Barret Y *et al.* Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* 23(2), 312–328 (2006).

4 Krouwer J. Estimating total analytical error and its sources: techniques to improve method evaluation. *Arch. Pathol. Lab. Med.* 116, 726–731 (1992)

5 Grant R, Hoofnagle A. From lost in translation to paradise found: enabling protein biomarker method transfer by mass spectrometry. *Clin. Chem.* 60(7) 941–944 (2014)

6 Jones B, Schultz G, Eckstein J, Ackermann B. Surrogate matrix and surrogate analyte approaches for definitive quantitation of endogenous biomolecules. *Bioanalysis* 4(19) 2343–2356 (2012)

7 Stevenson L, Purushothama S. Parallelism: considerations for the development, validation and implementation of PK and biomarker ligand-binding assays. *Bioanalysis* 6(2), 185–198 (2014)