White Paper

Theme: Best Practices for Bioanalytical Methods: Recommendations from the Global Bioanalysis Consortium

Guest Editors: Binodh DeSilva and Philip Timmerman

Large Molecule Run Acceptance: Recommendation for Best Practices and Harmonization from the Global Bioanalysis Consortium Harmonization Team

Marian Kelley,^{1,11} Christopher Beaver,² Lauren F. Stevenson,³ Ross Bamford,⁴ Paula Gegwich,⁵ Yamamoto Katsuhiko,⁶ Dongbei Li,⁷ Samantha Little,⁴ Arumugam Muruganandam,⁸ Daniela Stoellner,⁹ and Ravi Kumar Trivedi¹⁰

Received 10 July 2013; accepted 21 November 2013; published online 7 January 2014

Abstract. The L1 Global Harmonization Team provides recommendations specifically for run acceptance of ligand binding methods used in bioanalysis of macromolecules in support of pharmacokinetics. The team focused on standard curve calibrators and quality controls for use in both pre-study validation and in-study sample analysis, including their preparation and acceptance criteria. The team also considered standard curve editing and the concept of total error.

KEY WORDS: acceptance criteria; accuracy and precision (A&P); curve editing; quality controls; standard curve calibrators; total error.

INTRODUCTION

Of the 20 teams organized by the Global Bioanalytical Consortium, the L1 Team was charged with addressing the Large Molecule Run Acceptance criteria for standard curve calibrators and quality (validation) controls in validation and sample analysis. The membership included representatives worldwide, with participants from America and Canada, Europe, India, China, and Japan.

SCOPE

The scope of team L1, Large Molecule Run Acceptance, included the following topics:

- Preparation of standards and quality controls (QCs) for pre-study validation and in-study sample analysis
- Assessment of the nonlinear calibration model “goodness-of-fit” [accuracy (percent relative error) and imprecision (CV%)] for calibrators in pre-study validation and during sample analysis
- Considerations for standard curve editing and handling anchor (accessory) points
- Assessment of accuracy and precision of quality controls in pre-study validation and during in-study sample analysis
- Concept and application of total error

In general, the team's recommendations on ligand binding assay (LBA) acceptance criteria employed in both validation of ligand binding methods and during sample analysis are aligned well with relevant published regulatory guidance and white papers (1–8). While many of the countries represented on the Global Harmonization teams have published formal guidelines for regulating ethical drug development (9–17), to date, only a few (1,2) have specifically addressed the topic of large molecule bioanalysis. Throughout the course of deliberations and discussion, the team was mindful to ensure that any recommendations put forth were done so with clear scientific rationale. It is acknowledged that while many assessments should be considered, not all will be found to be necessary, nor appropriate, for every molecule in all situations. For example, the endogenous nature of some

¹ MKelley Consulting LLC, 1533 Glenmont Lane, West Chester Pennsylvania 19380, USA.

² inVentiv Health Clinical, 301D College Road East, Princeton, New Jersey 08540, USA.

³ Biogen Idec, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA.

⁴ Covance Laboratories Ltd, Otley Road, Harrogate, North Yorkshire HG3 1PY, UK.

⁵ Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, Ridgefield, Connecticut 06877, USA.

⁶ Kyowa Hakko Kirin, 9420 Athena Cir, La Jolla, California 92037, USA.

⁷ WuXi AppTec Biopharmaceuticals Co., Ltd., 288 FuTe ZhongLu WaiGaoQiao Free Trade Zone, Shanghai 2000131, China.

⁸ Affigen Biosolutions Pvt Ltd, 265/1F KSSIDC, Bommasandra Industrial Area, Bangalore, Karnataka, India, 560099.

⁹ Novartis Pharma AG, P.O. Box CH-4002, Basel Switzerland.

¹⁰ Syngene International Limited, Biocon-Bristol-Myers-Squibb Research Centre, Syngene International Ltd, Plot No. 2 & 3, Bommasandra IV Phase Jigani Link Road, Bangalore India, 560 100.

¹¹ To whom correspondence should be addressed. (e-mail: mmk48@Comcast.net)

biotherapeutics must be considered in the preparation of calibrators and controls. The application of rigorous scientific principles in determining the appropriate assessments to be performed should ensure the reliability of the assay and the data generated.

PREPARATION OF CALIBRATORS FOR VALIDATION AND SAMPLE ANALYSIS

The team noted that the practices for utilization of fresh *versus* frozen standards and QCs differed across the industry during both validation and sample analysis. It is understood that some researchers prefer to use standard curves consisting of frozen calibrators throughout validation while others advocate the use of fresh, arguing that the preparation of calibrators is a major contributor to robustness of the method (as discussed below). Whichever is used, the team strongly recommends consistency between validation and sample analysis practices, i.e., the validation should mirror how the sample analysis will be performed in production. If a method will be run during production with frozen calibrators, then the calibrators used in the validation should be analyzed similarly. For frozen standards and intermediate stock solutions, the team advocates verifying the calibrator concentrations upon preparation and then performing a stability run against freshly prepared calibrators at the end of validation. Since this approach may be perceived as performing a validation “at risk,” this approach should be supported by stability data (freeze-thaw, bench top, and long term) obtained during method development prior to validation. This *a priori* knowledge will permit the testing laboratory to enter the pre-study assay validation phase confident that sufficient stability will be established to cover the use of frozen aliquots during the accuracy and precision (A&P) validation runs. Likewise, if a sub-stock(s) of reference material is used over a period of time to prepare calibrators, the stability of these sub-stocks must also be confirmed. It may be argued that since a typical validation takes no more than 2 weeks to execute, the ability to establish 2 weeks' stability should be manageable.

The L1 Team was in agreement that the verification/qualification of calibrators prior to start of validation was a good scientific practice. The team felt that the method being validated consists primarily of the assay format and platform. Prior to moving into validation, establishing the appropriateness of the critical elements of the method (e.g., antibody pairs, appropriate dilution of detection reagent, and the accurate preparation of the calibrators and QC's) permits the validation to focus on the method, *per se*. There are many causes of variability which are reflected in A&P. This process helps to eliminate the contribution of unacceptably prepared calibrators and QCs which lead to run failures, not because of the inherent variability but due to inappropriately prepared calibrators.

Alternative approaches to the process of pre-qualifying calibrators and QCs have been brought to the team's attention. While we stand behind the scientific principles that guided the recommendation, the team acknowledges that there may still be dissenting opinions within the industry. An alternate approach is to use standards in the A&P validation runs without pre-qualification. The variability of the standard preparation can then be accounted for in the determination of

the method accuracy and precision, during A&P validation phase. Once this variability is assessed, bulk preparations may be used and pre-qualified to the acceptable criteria and used in the sample analysis phase of the method. It must be pointed out, however, that the variability demonstrated during validation will not be reflective of actual sample analysis.

The advantages of adopting the pre-qualification approach for these evaluations are as follows:

- It ensures that the validation data accurately reflect the level of variability to be expected during sample analysis
- The assessment targets primarily an assay's variability in relation to repeated assessment of the same samples multiple times, rather than the additional variability of differences in inter-individual spiking
- It mimics more closely the actual conditions of use under which the assay typically will be run
- For a substantial portion of assay work plans, particularly those used in high volume, high-throughput analyses, standards and QCs will be prepared, qualified, stored, and utilized for the duration of a study

PREPARATION OF VALIDATION SAMPLES/ QUALITY CONTROLS FOR VALIDATION AND SAMPLE ANALYSIS

Since QCs are intended to mimic samples, which will always be frozen, it is recommended that QCs be prepared, verified, and frozen to support pre-study validation and sample analysis. The team suggests that acceptance criteria for qualifying newly prepared QCs should be more stringent than for assay acceptance. For example, in an assay with run acceptance criteria of $\pm 20\%$, a high number of failures may occur if the acceptance for qualification approaches an accuracy of 20%. While no specific target is prescribed, the team felt that a preparation that did not deviate much beyond 10% would be a suitable target. Standards and QCs are typically qualified by independently preparing each from reference standard (drug substance) and demonstrating that (1) the standard curve meets proposed acceptance criteria and (2) that all QCs measured from the curve also pass criteria. A greater level of rigor may be introduced in the verification process by testing five levels of QCs (lower limit of quantification (LLOQ), low quality control (LQC), medium quality control (MQC), high quality control (HQC), and upper limit of quantification (ULOQ)) which will demonstrate that the entirety of the standard curve range has been cross-checked with independently prepared QCs.

ASSESSMENT OF CALIBRATORS DURING VALIDATION AND SAMPLE ANALYSIS

The objectives for assessment of standard curves differ between pre-study validation and in-study sample analysis. During validation, the goal is to verify the acceptability of a calibration model and weighting to support regulated bioanalysis. In contrast, the goal during sample analysis is to

confirm an acceptable standard curve during a run of test samples. Standard curve assessment must be performed independently of the acceptance of the quality controls in both pre-study validation and sample analysis and should precede the evaluation of the QC samples. The recommendations outlined in the guidance on Bioanalytical Method Validation for ligand binding assays regarding acceptance criteria for the standard curve during pre-study method validation from the EMA (1) and the FDA (3) and the DeSilva white paper (5) have been confirmed as recommendations by the L1 Team. Specifically, 75% of the calibrators in the curve (with a minimum of six valid standards within the anticipated calibration curve range and excluding anchor calibrators) must have back-fitted concentrations with a relative error of $\pm 20\%$ (versus the stated nominal concentrations) in order to pass acceptance criteria. In most ligand binding assays, as individual replicates (standards, QCs, and incurred samples) are typically assayed in duplicate or greater and the mean of the value is utilized, an additional criteria is that the mean %CV (between wells) ≤ 20 . It is important to note that the CV criteria for standards may vary depending on the software package used in the calculation. In some packages, the curve is fit to the mean of the raw values. In general, it is recommended that the standard curve contains greater than 6 standard points; anchor or accessory standard points outside of the analytical range are not included in this number. Masking or removal of standard points in order to obtain an acceptable standard curve is an acceptable practice providing the final standard curve, following the masking/removal exercise, still contains at least 75% of the original standard points (with a minimum of 6 valid points). When masking is employed, it is strongly recommended that an objective, step-by-step process be defined *a priori* and included in either a standard operating procedure covering method validation, a general acceptance criteria SOP or in the SOP for the specific test methodology, in order to avoid subjectivity in its application. The pattern of masking should be consistent across a pre-study validation and sample analysis studies.

In the majority of assays, samples (plus calibrators and QCs) are analyzed in duplicate. Particular caution should be exercised when allowing for partial masking of standard points within the curve (e.g., one well/value of a duplicate), particularly when standards are run in duplicate, as this can produce a degree of subjectivity, whether real or apparent, in standard curve reprocessing that may prompt scrutiny of the data. It is recommended that when calibrators are run in duplicate that both wells of the replicate be masked when reprocessing is allowed. In the rare cases when calibrators are run in $n > 2$, a standardized decision process should be utilized to identify which of the values comprising the replicate should be eliminated.

There was a discussion among the team on the decision flow of standard points editing. The approach of first masking points out-of-range for precision and then masking points out-of-range for accuracy by the order of significance is recommended. Significant levels or changes in the sequence of masking in a validation or sample analysis study may reveal a lack or a deterioration of method robustness that the testing laboratory should consider investigating. Extra care should be taken when 2 consecutive standard points are masked/removed, as this may change the standard curve

shape and impact QCs and sample calculation. In fact, some labs advocate not permitting masking 2 consecutive points.

The use of anchor points is a common practice in the development of LBA, and the number employed depends entirely on the assay being developed. In general, it is recommended to use the fewest needed to control the curve. Consensus was not reached on the acceptability of masking anchor or accessory standard points to generate an acceptable standard curve. Some testing laboratories never remove anchor/accessory standard points during their data analysis, while others do. It was agreed that, if the testing laboratory allows for the masking of anchor or accessory points, clearly defined, objective, step-by-step criteria by which this should be performed be defined to avoid subjectivity in the reprocessing, as is expected of standard curve calibrator points. At the completion of accuracy and precision runs, the cumulative (inter-run) performance of the standard curves should be assessed and documented.

During sample analysis, similar acceptance criteria for the standard curve as those during the pre-study validation were endorsed. That is, 75% of calibrators (resulting in a minimum of 6 points) must be within 20% of their nominal values (25% at the LLOQ standard and ULOQ standard).

It is important to clarify a difference between pre-study validation and sample analysis as relates to the handling of the LLOQ and ULOQ calibrators in curve reprocessing. During validation, validation samples are typically being run at five levels (LLOQ, LQC, MQC, HQC, and ULOQ) in order to establish the assay range through the performance of samples in terms of their accuracy and precision. It is not acceptable to mask the concomitant LLOQ or ULOQ standards in those runs as there would be no standard at one or either of those levels. Therefore, during validation, all precision and accuracy assays must contain LLOQ and ULOQ standard points and meet standard curve acceptance criteria. However, during sample analysis, the range and performance of the assay have been established during pre-validation. It is therefore acceptable to remove the LLOQ or ULOQ standard in the runs (providing the above criteria are maintained). Should either the LLOQ or ULOQ standard be removed from the curve, the next standard point (either up or down) will become the new LLOQ or ULOQ for the assay. A corollary of this practice is that all samples with a signal below the "new" LLOQ or above the "new" ULOQ of the assay will need to be repeated—even if their initially back-calculated concentrations were within the analytical range of the method established during pre-validation.

There was discussion around the use of duplicate values or mean values for generation of the standard curve. Some software products calculate a curve of best fit from single standard concentration values obtained from the mean of the two values for each replicate; other products have the option to produce the curve by keeping the duplicate concentration values of each standard separate and fitting the curve through the two values, rather than the mean. An informal evaluation of several data sets processed in both manners revealed no significant difference in terms of the standard curve regression and/or performance. Therefore, either of these processes is acceptable.

The team also considered the usefulness of using additional parameters such as curve parameters (e.g., slope,

midpoint in case of a four PL fit) for monitoring validity of the data as recommended by Findlay (7). A change of midpoint and slope over time could indicate degradation of reagents and/or reference standard. An anecdotal example was discussed within the team where the reference standard was degrading temporally. Since the QCs were degrading at the same rate, the QCs continued to “pass” and consequently the run was accepted. Monitoring the midpoint could have suggested an issue and led to an investigation. Although there is a clear benefit in monitoring long-term trends of curve parameters, it was not suggested to have this as requirement for run acceptance.

QUALITY CONTROLS (VALIDATION CONTROLS)

Quality controls, whether used for validation or sample analysis, should be prepared from drug product (or intermediate if a sub-stock is employed) independent of the calibrators' stock preparation, verified, and frozen to mimic the study samples. At a minimum, the QCs should be tested from the frozen state for the completion of the A&P. Some may argue that at the time of validation, stability may not be known, but typically, stability is verified during method development. Because reconstitution/dilution of reference material can be a source of variability, it is recommended to prepare and qualify batch preparations of controls. As stated above, the “acceptance” of the nominal concentration of these batches should be somewhat more rigorous than for method acceptance since variability conferred with the preparation is added to the variability found performing the assay.

Some *a priori* process should be devised for qualification of new batches of standards and QCs over the life of the assay. It is highly recommended that some longitudinal QCs are set aside to aid in bridging the lots of QCs.

ASSESSMENT OF QCs DURING VALIDATION

At least five QC concentrations should be used to determine accuracy, precision, and the total error of the method during validation (5,6). As a suggestion, the concentrations may be prepared at the anticipated LLOQ, ≤ 3 times the LLOQ, mid-range of the curve (e.g., near geometric mean), high (70–80% of the ULOQ), and anticipated ULOQ. Typically placing the LQC above the second calibrator and the HQC below the second highest calibrator can serve to protect a run (during sample analysis) should the LLOQ or ULOQ fail. For the assessment of accuracy and precision, measurements should be made with a minimum of three replicates of each concentration across at least six independent assay runs over several days (5). For assays that tend to be variable, more runs can be performed. The conduct of the validation should be representative of the actual study samples analysis, i.e., when a study sample result is reported as a mean of two replicates, then during validation, QC results should be the mean of two replicates (i.e., using two wells per QC sample result). If the study samples are expected to be run by more than one analyst, this fact should be taken into consideration during pre-study validation design, i.e., a fit-for-purpose risk-based approach.

Ideally, assay performance would be best reflected if all A&P runs are used (none rejected). Thus, it is suggested that all of the accuracy and precision runs be performed and assessed as a whole for inter- and intra-assay A&P and total error (TE).

However, if the QCs consistently fall outside the target acceptance criteria, the validation may need to be suspended to allow for an investigation. Depending on the outcome of the investigation, the validation may continue if no major change in the method has been implemented. If a major change to the method is implemented, the validation should start fresh.

Once the A&P have been calculated, method criteria can be set for the remainder of the validation. These criteria are used for assessing the remaining validation parameters, e.g., stability, dilutional linearity, *etc.*

During validation, the within-run and between-run accuracy and precision mean concentrations should be within 20% of the nominal value at each concentration level (25% at the LLOQ and ULOQ). The total error (i.e., sum of absolute value of the percent relative error and percent imprecision) should not exceed 30% (40% at LLOQ and ULOQ). Analyze all runs using an appropriate statistical method to determine both intra- and inter-assay precision and accuracy. (See the AAPS Biotech Section Discussion Board for one example of an A&P statistical package).

It should be noted that guidance documents allow for wider acceptance criteria, with supporting justification. Although 20% is the target acceptance criteria at the start of validation, the final acceptance criteria of a validated method should be data driven taking into consideration the intended use of the data. Final acceptance criteria of 25% may be acceptable in instances where validation data supports this. Alternately, a more robust method may support tighter criteria of 15%.

ASSESSMENT OF QCs DURING SAMPLE ANALYSIS

As already described during in-study sample analysis, the standard curve is assessed first for acceptance without reference to the quality controls. The study sample results are typically reported as a mean of two values; therefore, each QC result should also be reported as a mean of two back-calculated concentrations. Three concentrations (low, mid, and high) are assayed twice (four wells = two reported values per concentration) which are used to assess the validity of a run. Assessing the means, 4/6 QCs (67%) must be within 20% accuracy and precision (using the back-calculated concentrations and not the raw signal) and no two at the same concentration may be rejected. When more than two reported values are used in a validation, at least 50% must be acceptable at each level.

It is not uncommon for the drug concentrations to be very high for some biotherapeutics. A large dilution is, therefore, typically required to allow the assayed values to fall within the curve. In cases where the study is unblinded, for instances when the pharmacokinetics is known, it is acceptable to aim the dilution to the mid-part of the curve. In these cases, concentrating QCs in one narrow area is not recommended. The QCs should continue to be spread throughout the curve. This is very different from having a large subset of samples fall within a narrow range of a curve.

TOTAL ERROR

TE expresses the closeness of agreement between a measured test result and its theoretical true value by describing

the combination of systematic (bias, expressed as mean random error (RE)) and random error (imprecision, expressed as %CV) components (7) when assessing the overall inter-assay precision and accuracy. When calculating TE, the % CV does not refer to the duplicate wells of a single validation sample, but to the total variability of all validation samples assayed during the A&P portion of the validation. On the other hand, cumulative target acceptance for validation QCs during pre-study validation is typically $\pm 20\%$ RE and $\leq 20\%$ CV (which potentially would allow for up to 40% total error).

There is a disconnect then between these more liberal target acceptance criteria for pre-study method validation and the more stringent 4/6/20 acceptance (which translates into a TE of no more than 30%) for sample analysis where the 20% in the 4/6/20 rule possesses an inherent total error. The acceptance for QCs monitored during in-study sample analysis is based on a fixed interval of $\pm 20\%$ of accuracy (nominal) and %CV (between the two wells). Should a method *just* meet the validation criteria above, we would expect to see many failed runs during sample analysis when we apply the more stringent 4/6/20 rule. To bridge this disconnect and bring alignment between validation acceptance and in-study acceptance, De Silva *et al.* (5) recommended the use of total error as an additional criterion, i.e., if the method approaches both 20% accuracy and 20% precision (or a TE of 40%), one or both of those parameters must be improved to a point where they do not exceed 30%. During sample analysis, the 4/6/20 rule would continue to apply. If you are willing to accept a method with TE closer to the 30% or higher as valid, you will likely fail a large number of sample analysis runs. It will be necessary to scientifically defend why this method is acceptable for the purpose of that particular validation. This is a recommendation not only in the De Silva paper (5), but also in the most recent EMA guideline (1). It should be noted that mathematically the recommendations in these publications employ a surrogate computation for the calculation of total error. Alternate computational approaches are acceptable.

VALIDATION FAILURES

Until the A&P is fully calculated, assay runs during validation may only be failed if the standard curve fails. Once the A&P part of the validation is complete, the assay QCs (LQC, MQC, HQC) are used to determine the validity (or failure) of the other validation parameters. At this point, the TE should also be given some consideration prior to moving into sample analysis.

CONCLUSION

The team found that across the globe, the generally accepted run acceptance criteria found in many countries' guidelines and cited in several white papers are well aligned. This manuscript serves to reinforce those guidelines and perhaps provide further clarification especially concerning the concept of total error and how it is applied to method validation.

ACKNOWLEDGMENTS

The team would like to thank Drs. Ronald Bowsher and Binodh DeSilva for their contributions to the discussions on run acceptance.

REFERENCES

1. European Medicines Agency, Committee for Medicinal Products for Human Use. *Guideline on bioanalytical method validation*. July 2011.
2. Health Products and Food Branch. Guidance Document: Non-clinical laboratory study data supporting drug product applications and submissions: adherence to Good Laboratory Practice. Ottawa, ON, Canada (2010).
3. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine. *Guidance for industry: bioanalytical method validation*. Federal Register. 2001;66:28526–28527.
4. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) March 2003. <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm070124.pdf>.
5. DeSilva B, Smith W, Weiner R. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res*. 2003;20:1885–900.
6. Kelley M, DeSilva B. Key elements of bioanalytical method validation for macromolecules. *AAPS J*. 2007;9:E156–63.
7. Findlay JWA, Dillard RF. Appropriate calibration curve fitting in ligand binding assays. *AAPS J*. 2007;9(2):E260–7.
8. DeSilva B *et al.* 2012 white paper on recent issues in bioanalysis and regulatory agencies' alignment towards multiple harmonized bioanalytical guidance/guidelines. *Bioanalysis*. 2012;4(18):2213–26.
9. Findlay JW, Smith WC, Lee JW, Nordblom GD, Das I, *et al.* Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J of Pharm Biomed Anal*. 2000;21(6):1249–73.
10. OECD series on principles of good laboratory practice and compliance Monitoring. Number 1. OECD principles on good laboratory practice (as revised in 1997). 1–39 (1998).
11. Ministry of Health and Welfare (Japan) Notification No. 443, Guidance for toxicokinetics, July, 1996.
12. Japan Pharmaceutical Administration and Regulations. Information in English on Japanese regulatory affairs, March, 2012. <http://www.jpma.or.jp/english/parj/1003.html>.
13. Ministry of Health and Welfare (Japan) Ordinance No. 21, Ordinance on the GLP Standard for Conduct of Nonclinical Safety Studies of Drugs, March, 1997.
14. State Food and Drug Administration (China) Guide for the research of human bioavailability and bioequivalence about chemical drug, March, 2005.
15. Ministry of Health. National Agency for Sanitary Vigilance—Anvisa. Minimum requirements for the validation of bioanalytical methods used in studies with the purpose of registration and post-registration of medicines. RDC No. 27, May 2012.
16. Ministry of Health. National Agency for Sanitary Vigilance—Anvisa. Guide for the preparation of technical report on bioavailability/bioequivalence. RE No. 895, May 2003.
17. Australian Government, Department of Health and Ageing Therapeutic Goods Administration. Australian Regulatory Guidelines for Prescription Medicines, June 2004.