ICH M10 Guidelines Impact on Bioanalytical Method Development and Validation[i]

▷ **Chris Kafonek,** Manager, Method Development, Bioanalytical Sciences, Celerion

Introduction:

The final ICH M10 Guideline (M10) will be an important milestone in the evolution of bioanalytical method development, validation, and regulatory compliance. Compliance with bioanalytical validation guidelines is important for review and approval of new and generic drugs without significant and unnecessary delays to evaluate the quality of the bioanalytical data.

Similar to the way traffic laws establish expectations for drivers and facilitate their progress on the roads, guidelines for bioanalytical method validation establish expectations for method validation and facilitate review of bioanalytical data in drug applications. As drivers travel from state-to-state in the U.S., the traffic laws are substantially similar and this harmonization contributes to traveler safety and minimizes citations for violations. The M10 will create similarity between regulatory expectations in multiple regions, including the U.S., Europe and Japan, regarding bioanalytical method validation and acceptance criteria, facilitating drug approvals worldwide.

General Comments

In developing their individual guidelines, regulatory agencies have prioritized their recommendations for validation of bioanalytical methods to be used in the analysis of samples from bioequivalence studies where concentration data for the generic are compared to a reference product. As the approval of generics is based entirely on the statistical comparison of the concentration data in lieu of safety, tolerability, drug interaction, and other types of studies, the quality of the bioanalytical data are of critical importance.

For other types of studies, particularly early safety studies such as the single- and multiple-ascending-dose (SAD; MAD), the statement

"The objective of the validation of a bioanalytical assay is to demonstrate that it is suitable for its intended purpose."

may be the most important statement in the M10. This statement may make it easier to scientifically justify an accommodation to perfect compliance when application of the method is less critical than in bioequivalence studies. Although it may seem easier to be compliant with a definitive Guideline, a definitive Guideline does not absolve end users and bioanalytical scientists from their responsibility to evaluate each method with respect to its intended purpose.

Preparing for the analysis of samples from bioequivalence studies is fundamentally different than preparing for the analysis of samples from SAD or MAD studies. If an approved reference product is available, information on the expected maximum concentrations (Cmax), the intersubject range of Cmax values, and the elimination half-life are almost certainly available. It should be possible to select a concentration range for the assay that has a lower limit of quantification (LLOQ) sufficiently sensitive for the measurement of approximately five half-lives of elimination and an upper limit of quantification (ULOQ) that minimizes the number of samples that must be diluted into the calibration range.

During SAD and MAD studies, it is advantageous to have large calibration ranges, as the range of plasma concentrations is typically large when all cohorts (dose levels) are considered. It is desirable to have an LLOQ sensitive enough to measure concentrations beginning with the first "no adverse effect" or NOEL dose. Perfect compliance with a regulatory Guideline may not be necessary to obtain reliable data for dose escalation decisions or selection of the therapeutic dose.

For example, carryover that is significant when comparing the response in a blank following a ULOQ standard to the response of the LLOQ standard may not be significant with respect to the range of concentrations observed within a cohort of a SAD or MAD study – especially if the samples are analyzed in the sequence of collection. A batch that does not have a signal-to-noise ratio greater than 5 for the LLOQ may still be reliable for a cohort where nearly all of the concentrations are greater than the mid QC.

During the discussion at WRIB (Workshop on Recent Issues in Bioanalysis) in early April regarding the placement of the mid QC, regulators made it clear that the mid QC should be at 50% of the ULOQ. Several people pointed out that this could leave a relatively large range of concentrations unmonitored between the low and mid QCs. After much discussion between industry and regulatory representatives, it became clearer that this recommendation was based, at least in part, on an expectation of where the anticipated mean Cmax should fall in a calibration range tailored to a therapeutic dose in a bioequivalence study and reflected the critical importance of the Cmax in the statistical comparison for bioequivalence. For

other types of studies, the mid QC placement might be less critical and the geometric mean would be appropriate. The use of more than one QC between the low and high QCs would also be acceptable.

Bioanalytical methods developed and validated for SAD and MAD studies are not routinely (re)-optimized after selection of a therapeutic dose or range of doses, but an evaluation of the calibration range and the overall method performance between the SAD/MAD studies and the selection of the therapeutic dose should be routine. A method with a large range that has some carryover or borderline sensitivity is likely to cause unnecessary time and effort in routine analysis of samples from therapeutic dose studies. Truncating the range can significantly improve throughput by minimizing effort required to meet initial criteria for signal-to-noise ratio and carryover, as well as reducing re-injections and re-extractions. Poorly performing internal standards (ISTD) should be replaced with high quality stable-labeled ISTDs, possibly incorporating carbon-13 or nitrogen-15.

Method Development

The M10 stresses the importance of method development "to define the design, operating conditions, limitations and suitability of the method for its intended purpose and to ensure that the method is optimized for validation". It seems obvious that method development is necessary to establish a method that is capable of meeting validation acceptance criteria, but this purpose statement is clear that method development should go beyond determining validation capability.

From the perspective of a Contract Research Organization (CRO), the fulfillment of this objective requires open communication from sponsors. Sponsors looking for tandem mass spectrometric (LC-MS/MS) bioanalysis services often provide a method, typically using protein precipitation with a fast reversed-phase gradient that "worked great" for their animal samples. They want the CRO to "just validate it for human plasma" to save money and the time that thorough method development would require. Other sponsors want to develop the method and transfer it to the CRO, but for various reasons (e.g. differences in equipment, communication gaps, etc.) methods rarely transfer in a straightforward manner and much time and consultation is necessary to identify the critical differences. Even with appropriate confidentiality agreements in place, sponsors can be reluctant to share information about analyte structures and physicochemical properties (solubility, functional groups, pKa values, LogP or LogD values) that are critical in optimizing the bioanalytical method. Attempts to control the information about the analyte and the development of the method can be counter-productive to an optimized method. In addition, should the performance of the method deteriorate during the analysis of study samples, incomplete knowledge of the analyte can hinder problem-solving efforts and cause additional delays.

While bioanalytical method development does not require extensive record keeping, the M10 states "the applicant should record the changes to procedures as well as any issues and their resolutions to provide a rationale for any changes made to validated methods immediately prior to or in the course of analyzing study samples for pivotal studies." A CRO that adopts a method from another laboratory – sponsor or other CRO – is unlikely to know the history of the method development, the issues, or their resolutions, placing them at a disadvantage should investigation into the method performance during sample analysis be required.

Selectivity

While interference typically refers to a response that appears to result from the target analyte but does not, matrix effect is a change in the analyte response caused by a substance that does not produce a detectable response. Both interference and matrix effect are aspects of selectivity because they affect the response of the target analyte.

Bansal and DeStefano defined the Matrix Factor (MF) in 2007. The EMA specified preference for the MF test, the FDA did not specify a preferred method for the evaluation of matrix effect, and the M10 recommends evaluation of "3 replicates of low and high QCs, each prepared using matrix from at least 6 different sources/lots."

The detection of matrix effect is indirect – a decrease or increase in analyte response resulting from changes in the ionization efficiency, rather than from changes in extraction efficiency or injection volume, etc. The MF test evaluates matrix effects that co-elute with the analyte and/or ISTD. The matrix effect may be co-eluting because it is at the same retention time as the analyte in every extract or it may be more variable – appearing in some extracts but not others if it is "down-field" meaning that the matrix effect perpetrator is more highly retained and elutes in a subsequent injection.

If matrix effect is present during method development, e.g. from phospholipids, the method may be optimized to mitigate the impact of matrix effects; however concomitant medications also may be a source of matrix effect, and it is not realistic to anticipate during method development all of the compounds that could potentially be co-administered. For drugs commonly used as probes for impact on specific enzymes or transporters, e.g. itraconazole for cytochrome P450 3A4/5, or digoxin for P glycoprotein, it would be impossible to predict the investigative drugs that might be co-administered.

A review by Van Eeckhaut et.al. summarizes the most common approaches to the evaluation of matrix effect. There is certainly value in utilizing more than one type of evaluation during method development and tools that are not well suited to the application of validation acceptance criteria and generation of report tables (i.e. post-column infusion) may be used to investigate suspected changes in matrix effect at any time during method development, validation, or sample analysis.

Internal Standards

A change in matrix effect may be detected by monitoring the response of the ISTD – particularly stable-labeled ISTDs that co-elute with the analyte. Regulatory agencies have placed a great deal of importance on monitoring the response of the ISTD, but do not set acceptance criteria for the ISTD response. The industry should view this as an advantage because it allows them to establish criteria (or

no absolute criteria) for each individual assay. The ISTD variability is a characteristic of the assay. Routine monitoring of the ISTD response will reveal in a relatively short period whether the ISTD response is very consistent, somewhat consistent, or quite inconsistent. An assay that involves a solid-phase extraction, derivatization step, and liquid-liquid extraction may have a great deal of variability in the ISTD response, but that doesn't mean that it isn't a good method. The ability of the ISTD to track the analyte can be demonstrated by the improvement in the relative standard deviation (RSD) of the ratios compared to the RSD of the peak areas. A sudden change in ISTD behavior on a batch may be indicative of a change in the assay performance, while a sudden change in the ISTD response for unknown samples compared to calibration STDs and QCs may indicate a change in matrix effect. Both of these situations probably warrant investigation.

Analog ISTDs that do not co-elute with the analyte may not be indicative of, and are unlikely to compensate for, matrix effect that affects the analyte(s). In some cases, even a 2H isotopologue may be chromatographically separated from the unlabeled analyte. In other cases, such as polar analytes in urine, the ubiquity and severity of the matrix effects means that even a few hundredths of a minute difference in retention between a 2H isotopologue ISTD and the analyte can lead to inadequate compensation for the matrix effect. In these situations, a 13C or 15N isotopologue of the analyte would likely be a more effective ISTD.

Carryover

The M10 recommends that carryover be evaluated during validation by injection of a double blank sample following the ULOQ calibrator; analyte response in the blank should not be greater than 20% of the LLOQ calibrator and ISTD response should not be greater than 5% of the response for the ISTD.

The evaluation of carryover does not end with the validation, and in fact may vary within a batch, e.g. if a syringe plunger becomes loose during the injection of the batch. If the double blank sample for carryover evaluation does not meet the carryover acceptance criteria, the root cause may not be carryover, but it might be difficult to assess carryover impact based on the false positive carryover value.

Analyte susceptibility to carryover is not the same for all molecules. The use of various needle wash solutions can mitigate carryover in some cases, but some carryover in gradient methods may be "column carryover" and many autosamplers have a carryover limit of 0.1% in the specifications. If carryover cannot be sufficiently mitigated using rinse steps, other actions, such as truncating the calibration range or injecting blank samples after samples with expected high concentrations may be advisable.

Whole Blood Stability

The assessment of whole blood stability has sparked a great deal of discussion – including whether a separate exploratory method for the analyte in whole blood is required for the stability evaluation. The objective of this assessment is to demonstrate that the analyte is stable during the period from collection of the blood to harvesting of the plasma (or serum). Most problems encountered with this test arise from partitioning or equilibration artefacts as a result of spiking the analyte into whole blood. The heart is an incredible mixer and even intravenously administered drugs rapidly reach equilibrium between the intra- and extra- cellular spaces in the circulatory system. Drugs spiked into whole blood may not reach this equilibrium as rapidly, even with blood warmed to body temperature (although this facilitates the equilibration). Routine inclusion of an equilibrium test sample at 15 or 30 minutes during this stability assessment can usually distinguish between partitioning and stability issues. An analyte that demonstrates a difference between 0 and 30 minutes, but is stable between 30 and 150 minutes, likely just needed a little time to reach its partitioning equilibrium.

Analytes with pharmacological targets found primarily in white blood cells, such as zileuton, present a challenge in reproducibility, as it can be difficult to avoid the transfer of, or to transfer a consistent fraction of the white blood cells with the plasma. In such cases, an exploratory whole blood assay might be a more reproducible approach to this assessment.

Analytes that are also Endogenous Compounds

The M10 draft recommends that the biological matrix used to prepare calibration standards should be the same as the study samples and free of the endogenous analyte. Recognizing that it will rarely be possible to comply with both of these conditions, the M10 draft discusses four approaches to the analysis of endogenous compounds – standard addition, background subtraction, surrogate matrix, and surrogate analyte – without indicating a bias for or against any of the approaches.

1. Standard Addition – "every study sample is divided into aliquots of equal volume. All aliquots, but one, are separately spiked with known and varying amounts of the analyte standards to construct a calibration curve for every study sample."

   This approach is very labor intensive and Watson LIMS®, the most commonly used software in bioanalytical laboratories, does not accommodate this type of regression analysis for each study sample. The use of other software would require additional validation or intensive data review and verification.

2. Background Subtraction – "the endogenous background concentrations of analytes in a pooled/representative matrix are subtracted from the concentrations of the added standards; subsequently the subtracted concentrations are used to construct the calibration curve".

   This approach can be applicable if the response in a pooled/representative matrix is subtracted from the response in the fortified calibration standards prior to regression of the calibration curve. This approach requires manipulation (and verification) of the raw data prior to importing it to Watson LIMS®. It may also require significantly more effort to define a reliable LLOQ for the assay or to establish a relevant calibration range as the difference between the background response and the LLOQ must be large enough that the background is never greater than the LLOQ calibration standard.

3. Surrogate Matrix – "the matrix of the study samples is substituted by a surrogate matrix."

   This approach is not complicated with respect to the calculations and may be the only viable approach if the objective is to measure a decrease in the concentration of the endogenous analyte. The major challenge with this approach is demonstrating that differences in extraction efficiency and matrix effect on the response of the analyte in the surrogate and study matrices do not affect the accuracy of the quantification.

4. Surrogate Analyte – "stable-isotope labelled analytes are used as surrogate standards to construct the calibration curves for the quantification of endogenous analytes." A number of articles advocating this approach have been published but, as noted by the M10, "isotope standards may differ in retention time and MS sensitivity, therefore, before application of this approach, the ratio of the labelled to unlabeled analyte MS responses (i.e. the response factor) should be close to unity and constant over the entire calibration range. If the response factor does not comply with these requirements, it should be incorporated into the regression equation of the calibration curve."

   Again, the application of this approach becomes somewhat labor and calculation intensive as Watson LIMS® does not accommodate the calculation of sample concentrations directly from a stable-isotope labelled calibration curve. A correction for differences in response could be applied to the concentration after interpolation, but incorporation of the correction factor into the regression equation would be much more difficult. In addition, a second isotopologue distinguishable from both the analyte and the surrogate analyte, and that does not interfere with either, would be needed as an internal standard.

The surrogate matrix approach with QCs prepared in the same biological matrix as the study samples is the most straightforward approach. Extraction efficiency, matrix effect, and parallelism should be critically evaluated for endogenous analytes. Parallelism or dilution linearity is likely to identify significant matrix effect differences between samples in undiluted biological matrix and diluted with analyte-free surrogate matrix. The importance of a suitable stable-labelled ISTD for endogenous analyte assays can hardly be over-emphasized and 13C or 15N labelled isotopologues may be necessary, as they are much more likely to co-elute very closely with the unlabeled analyte and therefore better compensate for matrix effects.

Conclusions

It will never be possible for the regulatory agencies to write a guideline that sets out parameters and acceptance criteria for every method in every application. Scientific judgement and contemporaneous documentation of the rationale for decisions made with respect to testing performed and data accepted or rejected will always be important in reconstructing a study and evaluating the adequacy of the method validation.

The performance of ISTDs is critical to the long-term reliable performance of bioanalytical methods. The better an internal standard mimics the analyte with respect to extraction efficiency and chromatographic retention, the less likely the method will fail because of changes in matrix effect, extraction efficiency, or other issues such as variable injection volume or variable extract volume in normal phase methods where the extract can evaporate quite rapidly. The increasing use of ultra-high performance liquid chromatography (UHPLC) and the narrow peaks that are characteristic of UHPLC, place additional importance on the similarity in chromatographic behavior of the ISTD and the analyte. A difference in retention of 0.02 minutes on peaks that are 0.1 minute in width at the baseline can make a significant difference in response if the peaks elute in an area of rapid change in ionization conditions (i.e. on the slope of a suppression or enhancement).

The quantification of endogenous analytes might be improved with 21-CFR Part 11-compliant commercial software that was adaptable to standard addition, surrogate analyte, or background (response) subtraction approaches without manual manipulation and verification of the raw data. At the very least, software that could process data from all 4 approaches would permit comparison of the approaches and selection of the best approach for scientific reasons, rather than because of process-limitations.

References

I. ICH Harmonized Guideline, Bioanalytical Method Validation, M10, Draft version, Endorsed 26 Feb-2019

II. Bansal S, DeStefano A, The AAPS Journal, 9(1) Article 11 (2007)

III. Van Eeckhaut A, Lanckmans K, Sarre S, Smolders I, Michotte Y, J Chromatogr B, 877:2198-2207 (2009)