17 Analytical Method Development and Validation in Accordance to the Regulatory Guidelines

STEVE E. UNGER
Worldwide Clinical Trials Drug Development Solutions - Bioanalytical Sciences, Austin, TX, USA

17.1 SUMMARY

Present requirements to perform bioanalysis for submission to regulatory agencies have evolved as the result of bioanalytical workshops held in 1990, 2000, and 2006. Following the second meeting, the FDA issued its guidance in 2001. The European

Encyclopedia of Drug Metabolism and Interactions, 6-Volume Set, First Edition.
Edited by Alexander V. Lyubimov.

Medicines Agency (EMEA) has issued draft guidance, which will be issued as final guidance in February 2012.

Setting the assay range includes assessment of the lower limit of quantification (LLOQ), which is related to precision and accuracy of the lowest standard, as determined in replicate measurements, and is required to be within 20%. Chromatographic assays generally yield linear or quadratic response over a broad range of concentrations. Multiple days of validation data are needed to establish reproducibility of the regression model. The correlation coefficient of the standard curve must be >0.99. When validating in multiple species, a mixture of full and partial validations is common. As each matrix is unique, full validations should be undertaken whenever the matrix is changed.

For regulatory bioanalysis, matrix of the standards must match study samples. Included with each analytical batch are blank matrix, a sample with only internal standard (IS), and a minimum of six nonzero calibration standards. The absolute difference between the back-calculated and nominal concentration for each calibration standard should be within 15% for all standards, except at the LLOQ where $\leq$20% is acceptable. A minimum of 75% of the standards should be within these limits. Values falling outside these limits can be discarded, starting with the most deviant value and iteratively reregressing until all results are acceptable or the run fails. In cases where it is not possible to obtain the matrix in sufficient quantity to prepare standard and QC (quality control) samples, a surrogate matrix may be used. While it is preferable to have a certificate of analysis, some compounds available in very small amounts require only less qualification.

QC samples should be positioned evenly through the batch, reflecting the concentration of study samples. QC samples measured at a minimum of three concentrations in duplicate (or at least 5% of the unknown samples) are required. Spacing is within $3\times$ LLOQ, one near the center and one near the upper boundary (top quartile). Wider-range assays should include the addition of more QC levels to better represent study data. At least two-thirds overall and 50% at each level must be within 15% of their nominal value. When performing an assay validation, more replicates and additional QC samples are required to assess accuracy and precision at the LLOQ and to test dilutional linearity. A minimum of five determinations at each QC concentration are required. Additional QC samples include the LLOQ QC whose intraday and interday precision and accuracy needs to be within 20%. Dilution QC samples should test the maximum range over which the dilution is anticipated.

Matrix effect is the suppression or enhancement of ionization of analytes by components in the biological samples and is measured by comparing the response from a sample of blank matrix to which an equivalent amount of analyte was added after extraction to an equivalent amount in mobile phase. ISs in MS (mass spectrometry)-based assays can minimize the impact of matrix effects by normalizing response ratios. Having a highly variable matrix effect in individual subjects may cause irreproducibility. Absolute and IS-normalized matrix effects for six individual matrix lots are tested, and variability should be within 15%. When using stable isotope IS, this determination is unnecessary.

To accurately assess the recovery of analyte or IS from the sample matrix, response from an extracted sample is compared to a sample of blank matrix to which an equivalent amount of analyte was added after extraction. Apart from the criteria that recovery of analyte and IS be consistent, precise, and reproducible, there is no required minimum.

Carryover in an HPLC system is contamination by the preceding sample. If carryover is more than 20% LLOQ, its acceptance should be justified and appropriate precautions used to minimize its impact. In contrast to carryover, contamination is unpredictable and may occur during any part of the study (dosing, sample collection, or analysis).

Guidance does not mandate that within a study dilution QC samples are used, provided dilutions are conducted with like matrix and tested within validation. A preferable approach is to include dilution QC samples whenever study samples require it.

Procedures must evaluate the stability of analytes during sample collection and handling. The effect of different anticoagulants on stability should also be tested. Use of selective inhibitors is preferred since these can be added in smaller quantities and can result in minimum change to the matrix. Nonspecific binding of drug to containers also needs to be considered. In addition to stock solution stability, assessments in the unaltered matrix of qualifying QC samples include long-term stability under storage conditions, freeze–thaw cycles, bench-top (room temperature and/or ice processing), processed sample or extract stability, and reinjection integrity.

Specificity should ensure that background response from six independent lots of blank matrix is $<20\%$ of the response from the LLOQ sample and $<5\%$ of the IS. Selectivity is defined as "the ability of the bioanalytical method to measure and differentiate the analytes in the presence of components that may be expected to be present." Guidance requires that "each blank sample should be tested for interference, and selectivity should be ensured at the LLOQ." An assay should be tested whenever clinical studies are performed, where there is the potential for concomitant medication or when used in drug–drug interaction (DDI) studies.

Only qualified and properly maintained instruments should be used for drug analysis. To qualify instruments at time of use, a system suitability test may be used. This generally includes replicate testing of the analyte for sensitivity and precision near the LLOQ before starting an analytical run. Similar tests following the run may be used to confirm instrument stability.

A means to judge individual results, based on the response of ISs, is often used. Rules must be in place that allow *a priori* decisions to reject sample results that show aberrant IS response. Use of a stable label rather than a structural analog as IS is always preferable, as, once equilibrated within the matrix, it compensates for any changes.

Samples may be reanalyzed for an assignable cause such as a technical error, poor chromatography, unacceptable IS response, or evidence of carryover. The original result is discarded, and the sample is reanalyzed in singlet. The acceptable reanalysis value is reported. Samples may also be reanalyzed for anomalous results or for pharmacokinetic (PK) reasons. Typically, confirmatory reanalysis is performed in replicate and the original result is either confirmed or replaced.

There are a number of situations where the performance of QC samples may not adequately mimic that of study samples from dosed subjects (incurred samples). Recent studies have required the assessment of incurred sample reanalysis (ISR). Randomly selected subjects include a minimum of 10% of the total analyzed samples. For larger studies, 5% of the total sample size is allowed. The difference between the initial and reanalyzed values should be $\leq 20\%$ (small molecules) in at least 67% of reassayed samples. Unlike repeat analysis for anomalous results or for PK reasons, ISR is performed in singlet from selected samples about the C_{max} and in the elimination phase.

Simultaneously measuring multiple analytes increases the statistical probability of run failures. Results should not be rejected for the remaining analytes if one analyte

fails. Concentrations from the first accepted run should be reported. If this analyte is repeated in simultaneous assays when analyzing for different analytes, it is not necessary to quantitate the already reported analytes.

Tiered assays can be used when reference standards of metabolites are unavailable and a preliminary assessment of human exposure and safety coverage is needed. There are multiple strategies that have been used to allow this across-species comparison. Once the metabolism is fully known, a solid decision can be made regarding what regulated assay(s) would be effective in supporting later clinical drug development.

Assay transfer within a laboratory or across laboratories involves an assessment of its equivalence following the change. Depending on the nature of the change, this may involve either a partial or full validation. Conformance testing by analysis of QC samples is a good practice but not required. Cross-validation by testing study samples is unnecessary unless two different assays are used within a single study.

Proper installation and operational and performance qualification of all instruments is required before performing regulated bioanalysis. User specifications and a validation plan should be developed. System functionality can be verified and any gaps identified. User acceptance testing (UAT) should include a traceability matrix to regulations.

When unsure of clinical success, exploratory investigational new drug (eIND) studies allow the assessment of new chemical entities that have undergone limited safety testing. Provided there is a sensitive LC-MS (liquid chromatography-mass spectrometry) assay or a method such as AMS (accelerator mass spectrometry) is used, a subtherapeutic dose may afford the opportunity to quickly determine adsorption, distribution, metabolism, and excretion (ADME) properties.

Peptides, small proteins, and oligonucleotides are representative of mid-size molecules that can be directly measured using LC-MS. Larger proteins may be digested and indirectly measured by LC-MS analysis of a peptide surrogate. When using a mixture of ligand-binding and LC-MS assays, cross-validation using both QC and study samples can help to illustrate assay differences.

Dried blood spots (DBSs) offer numerous advantages, particularly from the in-life study. Bioanalytical challenges include recovering the analyte from filter paper without the introduction of matrix effects and overcoming sensitivity limitations for highly potent drugs. As more investigators explore the potential of DBSs in clinical studies, there will undoubtedly be regulatory bioanalytical guidance that comes forth.

The nature of regulated bioanalysis is ever changing. New methodology, technology, and novel drug candidates challenge past practices. Common procedures are now in place for validation and sample analysis of both large and small molecules. Drug development has integrated new cultures and perspectives into our science, making it essential that global regulatory requirements are standardized to allow for new drug submissions in all countries.

17.2 HISTORY OF REGULATED BIOANALYSIS

The term *bioanalysis* was derived in the 1970s to describe the quantification of drugs in biological fluids in support of PK studies. It evolved over the next decade to include toxicokinetics, the assessment of systemic exposures in animals undergoing toxicology studies. As many drug candidates show nonproportional dose–exposure relationships,

underpinning toxicity assessments with plasma or target-organ drug levels was a significant advance.

Problems related to preclinical research studies revealed serious issues with the conduct of safety studies submitted to the FDA. As a result, FDA published in December 1978 its good laboratory practices (GLP)s regulations. These regulations were collected in Title 21 "Food and Drugs" of the Code of Federal Regulations (CFR) as Part 58 "Good Laboratory Practice for Nonclinical Laboratory Studies" and applied to all nonclinical safety studies intended to support research permits or marketing authorizations of products regulated by FDA. For clinical *in vivo* bioavailability or bioequivalence study, 21 CRF 320.29 can be applied.

In Europe, adherence to GLPs is governed by the European Union (EU) law and, as in the United States, is applied to nonclinical safety testing. In 1981, the Organization for Economic Co-operation and Development (OECD) finalized its GLP principles. Later, the European Community adopted the OECD principles of GLP, requiring states to incorporate into their laws that nonclinical safety studies be conducted to GLP standards. ICH S3A provides limited bioanalytical guidance in the assessment of toxicokinetics, whereas ICH E6 requires documented competence in bioanalysis for clinical studies. EMEA CPMP/EWP/QWP/1401/98 states that the bioanalytical part of bioequivalence trials should be conducted according to GLP principles.

Although clinical bioanalysis is generally performed to the same high standard as nonclinical bioanalysis, it cannot be considered GLP. Methods are validated to the same standards, and there is adherence to many parts of the quality system that are fundamental to GLP regulations. The European Bioanalytical Forum has promoted the use of "regulated bioanalysis" to encompass studies that adhere to common regulatory guidance for bioanalysis rather than "GLP bioanalysis." Both terms will be found within the literature.

Present US requirements to perform bioanalysis for submission to regulatory agencies have evolved over the last two decades. Karnes *et al*. [1] provided early bioanalytical guidance. Shah *et al*. [2] published the conference report from the first bioanalytical method workshop. This report served as the foundation for bioanalysis for several years. The workshop defined accuracy, precision, selectivity, sensitivity, reproducibility, limit of quantification, stability, and recovery. Of these, the assessment of accuracy and precision was one of the most criticized parameters [3,4]. Statistical assessments from larger data sets were proposed as an alternative to fixed acceptance rules [5]. These arguments also found favor within the large-molecule bioanalysis community where less precise and accurate immunoassays had difficulty meeting rigid requirements. However, having simple, *a priori* rules by which to judge an assay validation and its application was welcomed by those performing small-molecule bioanalysis and was readily adopted.

During this period, two other guidance documents emerged from the International Conference on Harmonization (ICH), which were approved by the EU, the United States, and Japan [6,7]. While both dealt with analytical methods for pharmaceutical products, they nonetheless provided guidance on method validation. Because the workshop report was not an official document, the FDA decided that a guidance should be developed. The draft guidance was issued in January 1999. The second bioanalytical method workshop was held in January 2000 and reported later that year [8]. In 2001, the FDA issued its final guidance on bioanalytical method validation [9]. This guidance has served the small-molecule bioanalysis community for many years.

The large-molecule (biologics) community worked from separate guidance [10,11]. It was not until the third bioanalytical method workshop in 2006 that both the small- and large-molecule communities accepted similar rules for method validation and sample analysis. Both assess accuracy and precision in a similar manner; however, acceptance criteria differ. For large molecules, accuracy and precision were both required to be within 20% or 25% at the LLOQ, whereas the small-molecule community continued to require 15% (20% at LLOQ). Having both small- and large-molecule groups define common procedures and similar acceptance criteria allowed easier assessment of data by regulatory agencies.

The third bioanalytical workshop report was issued in 2007 [12]. Apart from the inclusion of the ligand binding or large-molecule bioanalysis into a common community, the most notable change was the requirement to demonstrate repeatability from incurred (study) samples. This had been a requirement of the Health Protection Branch for Canadian submissions from 1992 until the Therapeutics Products Directorate of Health Canada revoked the policy in 2003. Under this Canadian requirement, 15% of samples were randomly selected and reanalyzed.

Since 2007, ISR has become common practice for US submissions, which has added value by uncovering errors unnoticed from QC samples [13]. Other additions included an assessment of matrix effects [14] and carryover, as well as clarifications on stability testing and use of system suitability samples. More details on documentation were also added in the conference report, aiding in transparency and giving guidance on reporting failed runs.

Overall, there were limited changes to the prior procedures and acceptance criteria for small-molecule bioanalysis. There were significant changes for ligand-binding assays of large molecules. The report recognized the inherent differences in drug product manufactured from a biological process from drug substance generated by chemical synthesis. The heterogeneity of a biologic product, often including posttranslational glycosylation or phosphorylation, requires special consideration when defining validation criteria. Ligand-binding assays are often necessary to define not just chemical stability but the biological integrity in a stability assessment. In contrast, chemical stability of small molecules is more easily determined. Reagents, especially antibodies, used in ligand-binding assays may vary more than chemicals used in small-molecule assays. For these reasons, acceptance criteria were lessened. Assessment of total error (sum of the absolute value of mean bias and precision) was required to be within 30% (40% at LLOQ). No similar requirement of total error exists for small-molecule guidance.

The EMEA issued draft guidance on the validation of bioanalytical methods in September 2009, and requests for comments ended May 2010 [15]. Within it, many of the procedures and acceptance criteria within the FDA guidance and bioanalytical method workshop reports have been adopted. However, there are some notable differences. The EMEA guidance goes further in defining when the possibility of back-conversion of metabolites to parent should be tested. More detailed procedures on assessment of the matrix effect, particularly with regard to population diversity, have also been recommended. Inclusion of testing lipemic and hemolyzed plasma in the EMEA draft guidance parallels that found within the Brazilian regulations [16]. As a result of divergence from a commonly accepted FDA guidance, pharmaceutical companies are requesting that a common standard be adopted for global regulatory submissions [17,18]. Even minor differences in procedures or acceptance criteria could make a submission unsuitable in some countries. Therefore, establishing common

global expectations for method validation and sample analysis is critical to global submissions.

This chapter discusses regulatory bioanalytical requirements in a manner that attempts to provide guidance on how to achieve acceptance criteria. Deploying assays without fully characterizing their limits can result in failure runs. This is most apparent when assays are transferred to new laboratories or are used by new analysts. Best practices often include a standardized design and testing of assays during method development to better ensure that they are relatively insensitive to small changes. Knowing critical limits helps to hold an assay in control or provides the information needed to quickly diagnose and correct failures. We focus on the use of LC-MS, particularly in combination with mass spectrometry/mass spectrometry (MS/MS), to measure drug molecules and their metabolites to regulatory standards.

17.3 RESPONSE CALIBRATION

Sensitivity is defined as *the slope of the calibration model*. It should not be confused with the ability to detect or quantify an analyte in a given matrix. The limit of detection (LOD) is related to the signal-to-noise (S/N) ratio and, at a minimum, is required to be ≥ 3. The exact assessment may vary dependent on whether peak-to-peak or root mean squared (RMS) noise is used. The LOD is rarely used in the field of bioanalysis, except that chromatographic peaks of standards must have a signal that is fivefold greater than the background [9]. Some have proposed that the LLOQ should require a S/N ratio of ≥ 10. This is good guidance, as it is unlikely that all samples will be as free from interference as those analyzed during the validation. The LLOQ is related to precision and accuracy of the lowest standard, as determined in replicate measurements, and is required to be within 20%.

Having an understanding of what limits quantification at the extremes of the standard curve can guide method development and allow proper decision on assay range. In an ideal situation, detector response is proportional to concentration over a wide range. Provided there are limited losses due to adsorption (nonspecific binding), testing the assay range using pure solutions can establish the optimum detection range. Under these conditions, there are no recovery, stability, or matrix effects that can either dilute analyte concentration or its response. For a pure solution near the LLOQ, electronic noise limits detection. A major objective of method development should be the extraction and separation of drug from plasma to yield response limited by ion formation, transmission, and detection. Release of adsorbed analyte from a prior injection (injector carryover) can result in greater than the proportional response at low levels, so attention should be paid to the normalized response (response factors) to make a proper assessment. Near the upper limit of quantification (ULOQ), ionization and detector saturation can result in lower than the proportional response. MS systems using pulse counting electronics may also suffer from the inability to discriminate multiple ion events within the detection time frame. Within a 10-ns counting window, ion events $>10^8$ counts per second result in detector saturation. Analog electron multipliers do not suffer from this problem but are more limited by electronic noise at the LLOQ. With present instrumentation, ionization saturation generally occurs before detector saturation. This is typical for electrospray ionization (ESI) and is discussed later in

greater detail. Unless limited by adsorption, the assay range will be no greater when the analyte is placed into the matrix than when it is in a pure solution.

When spiked into matrix, chemical background often interferes with detection near the LLOQ. Less obvious background such as phospholipids in plasma, bile acids in bile, and salts or urinary acids in urine can modulate MS detection and are referred to as *matrix effects* [14]. Both chemical background and matrix effects can significantly impact measurements at low levels. While it may be easy to detect analyte at low levels, precise and accurate quantification is more difficult. When coeluted with analyte, the numerous and abundant components within biological fluids and tissues, compete for ionization with the analyte and affect response. If they vary significantly from the matrix used to prepare the standard curve, there is the potential for error. Use of a stable label as an IS compensates for variation in matrix effects and is a good addition to any assay. For multiplex assays, obtaining multiple stable label ISs can be difficult. Caution should be used when measuring parent drug and metabolites with only stable label IS of parent drug. High levels of parent drug can suppress ionization of its stable label IS, which, if also used to measure metabolites, can result in errors.

Unlike immunoassays, chromatographic assays generally yield linear or quadratic response over a broad range of concentrations. Method development should strive to establish the simplest fit, a linear fit with minimum weighting. Validation should strive to establish the best and most reproducible fit. The regression model must be justified and shown to be reproducible. Justification involves testing multiple regression models and different fitting parameters. These tests should use validated software applications, including the MS or LIMS (laboratory information management system) software, which will perform the sample analysis.

Linear fits are common; however, quadratic models can be used. A quadratic regression model can be used to address less than proportional response at increasing concentrations due to ionization or detector saturation. When using quadratic curves, a stable label IS is highly recommended to overcome intersample variations and their potential impact on MS response at its ionization limit. By using a stable label IS, the response ratio is better normalized, resulting in an extended dynamic range. Less than proportional response at decreasing concentration is often due to adsorption losses and associated changes in extraction recovery. This is common for hydrophobic drugs in matrices such as cerebral spinal fluid (CSF) and urine. Quadratic fitting should not be used to correct for nonlinear response at lower concentrations. Less than proportional response at low concentrations should be corrected by selecting a different container, adjusting pH to enhance solubility, or adding a solubilizing agent such as a surfactant or organic diluent. In both extremes, sensitivity is decreased as either the recovery near the LLOQ or detection at the ULOQ is impacted.

A much preferred approach is to require linear fitting of calibration curves. Linear fits can restrict the dynamic range that may require more dilution of study samples. However, a linear response model generally has better interday reproducibility. A linear fit is therefore the preferred and the anticipated response function. Deviations from a linear regression model may be challenged, particularly when fitting parameters show significant interday variations.

Weighting ($1/x$ or $1/x^2$) should be used for heteroscedastic data or whenever error is proportional to sample concentration. These corrections should be carefully scrutinized as daily changes in carryover or contamination can impact whether $1/x$ or $1/x^2$

weighting should be used. The power for weights approach evaluates the linear dependence of the logarithm of the standard deviations of peak response on the logarithm of sample concentration [19]. Other statistical approaches can also afford an unbiased selection.

Guidance requires the correlation coefficient (r^2) to be >0.99. It is likely that many regression models will meet the guidance criteria. Examining residual errors from validation data can aid in selecting the best fit. Proper documentation of the rationale for selecting the model is critical. Multiple days of validation data are needed to establish reproducibility of the regression model.

A common mistake is to define a linear model during validation only to find, following an improvement in instrument performance, that a quadratic fit may have been better. It is possible to inject less extract or simply retune MS performance to its less sensitive, linear state. However, if a wide assay range was needed, a quadratic fit may have been a better fit. When sensitivity is decreased, a quadratic fit is reduced to a linear fit (quadratic term $\rightarrow 0$). The reverse is not possible. Linear fits are preferred and can yield assays with reasonable dynamic range, provided extract purity and separation avoids ionization saturation. It should be noted that during the establishment of the regression model, variables, including injection volume, should be controlled. It is the total amount of coeluting components that affect MS response, so changes in their absolute amount will affect ionization saturation [20].

When validating in multiple species, a mixture of full and partial validations is common. Many analysts perform full (three days or more) validations in man and the primary (rodent and nonrodent) toxicology species with partial (one day or more) validations in other species. For instance, plasma assays in rat and either dog or monkey used for long-term (definitive) toxicology studies would receive full validation. Plasma assays for mouse (carcinogenicity or pharmacology), rabbit (reproductive), and dog (safety pharmacology; if monkey is used as primary) may be performed as partial validations. Others perform full validations in all species. Unless there is a reason to suspect a gender, or population difference, validations in humans generally assume equivalence. In contrast, validations in animals are generally both species and strain-specific. Testing any differences can be performed within the full validation to demonstrate equivalence on substitution. Changes in sample anticoagulant are not assumed to be equivalent and require at least a partial validation.

Provided there are no significant differences across species, the assay range established from the full validations should be applicable to other species. This simple assumption fails when there are significant interspecies differences in sample volume, drug potency, PK properties, stability, recovery, matrix effects, and endogenous interference. It is, however, preferred to test assay applicability across species before undertaking formal validations. Having a common analytical procedure reduces complexity in drug development. Much can also be learned from discovery studies, as assays were generally applied to numerous species and matrices.

Quantification of multiple analytes (comedications, metabolites) across species multiplies challenges in defining the proper assay ranges and regression models. Species differences in metabolism are common, resulting in different ranges and regression models for parent and metabolite. Multiple component assays should therefore be undertaken with great care to avoid deploying an assay with high failure rates. Some sense of assay requirements in each species can generally be gained by reviewing prior discovery studies, and this is helpful in devising a common analytical strategy.

The analyst's ability to control both contamination and carryover will also affect decisions on the LLOQ and assay range. Bioanalytical guidance restricts contamination of blank samples to within 20% of the LLOQ. EMEA also has guidance on measuring drug levels in control animals from toxicology studies [21]. Since the bioanalyst's role is to define whether control animals were contaminated, careful consideration of the LLOQ should be made to avoid false-positive detection and quantification. Contamination, held in control by more experienced analysts, may be increased when less experienced analysts undertake the analysis. Automation and proper training serve to reduce this risk.

The regulatory expectation is that assay range should be adjusted to study needs. It is common to have a high sensitivity, wide range assay during first-in-human studies. By the end of the multiple ascending dose (MAD) study, some indication of a safe, therapeutic dose should have been established. The analyst may therefore wish, at this time, to consider whether the assay range should be adjusted. Apart from being inefficient, dilution of study samples can amplify errors. This practice is more common to highly sensitive, narrow range immunoassays. However, dilution factors as great as 1000-fold have been used when measuring wide therapeutic margin drugs in high dose toxicology studies by LC-MS. QC samples must reflect study samples, so readjusting the range or adding extra QC samples may be needed.

As each matrix is unique, full validations should be undertaken whenever the matrix is changed. When urine data is critical to interpret PK or safety data, a full validation is needed and measurements may be required in all samples. For instance, in cases of compounds that show moderate-to-high renal clearance or where the kidney is the target site of action or toxicity, an assay should be fully validated in urine and a routine urine analysis may be required. Full validations are less needed to make early clinical assessments, particularly when animal studies show little parent drug is excreted in urine. Estimating urine levels at the highest single dose before deciding on the urine assay range for multiple dose studies is one means to better judge its requirements before validation. In either case, sufficient testing must ensure that sample collection avoids losses due to instability or irreversible adsorption. Measuring renal clearance and testing its dose proportionality at steady state is generally an objective that will require a validated assay. Most would accomplish this objective within the MAD study.

There is less guidance when demonstrating penetration into remote compartments. When critical to understanding PK–PD (pharmacodynamics) relationships, a full validation should be undertaken. For instance, measuring levels in synovial fluid for arthritis [22], bone for osteoporosis [23], CSF for CNS [24], or tumor for oncology [25], drugs would support target engagement. Tissue analysis is common in discovery to ensure penetration at the site of action, but translation to human studies can be difficult. There are many practical limitations, including obtaining tissue or bone samples, finding sources or surrogates for these matrices, assessing consistency in recovery of standards versus study samples, and ensuring the proper preparation of homogenates. Within toxicology studies, staff is often skilled and can prepare homogenates for storage before shipment and bioanalysis. Storing homogenates rather than tissue allows one to conduct a proper stability campaign. In contrast, preparing homogenates at multiple clinical sites is a challenge. These objectives should be supported using a full validation.

If one wants to demonstrate penetration to numerous clinical tissues but in a limited numbers of samples, a partial validation is generally acceptable. This would be the

case of an antibiotic tissue distribution study to determine whether numerous nonsystemic levels exceeded drug potency throughout the dosing interval (e.g., $C_{\text{tissue}} > \text{IC}_{50}$). A similar penetration assessment of antiviral agents measures their intracellular drug levels in peripheral blood mononuclear cells (PBMCs). Intracellular determinations afford a better understanding of drug uptake, efflux, and intracellular metabolism [26,27]. Since only a single matrix is involved and there are no limitations on its availability, a full validation would be expected.

17.3.1 Standards

For regulatory bioanalysis, standards must match study samples, which are most frequently plasma, blood, or urine. It is unacceptable to quantify study samples using another matrix standard curve. It is acceptable to dilute study samples within another matrix, provided standards are prepared in an identical manner. This may aid quantification by using a mixed calibrator curve or by diluting samples in a less problematic matrix.

Included with each analytical batch are blank matrix (sample without IS), zero standard (sample with IS), and a minimum of six nonzero calibration standards. The absolute difference between the back-calculated concentration and nominal concentration of each calibration standard should be within 15% for all standards, except at the LLOQ where $\leq$20% is acceptable. A minimum of 75% of the standards should be within these limits. Values falling outside these limits can be discarded, starting with the most deviant value and iteratively regressing until all results are acceptable or the run fails. When performing sample analysis, the exclusion of calibration standards cannot change the established model.

One standard curve is acceptable. Some prefer to use a second curve, often placing one at the front and the other at the end of the run. A significant change in sensitivity over the course of the run or a mistake in preparation will be noted by this approach. All results must be integrated into a common standard curve. Runs where the sensitivity changes over the course of the run by more than 25% will generally fail to match standards with QC samples.

It may not be possible to obtain the matrix in sufficient quantity to prepare standard and QC samples. In these instances, a surrogate matrix may be necessary. This is often the case in pediatric or human penetration studies. If unavailable, artificial CSF may serve as surrogate matrix for human CSF. Likewise, the same animal tissue from a closely related species may serve as a surrogate for human. A partial test of equivalence is always needed. Using predose or control matrix to prepare standard and QC samples is sometimes possible. Dilution using a more common matrix also helps to reduce the need for large quantities of matrix.

Although it is preferable to have a certificate of analysis for all analytes, some compounds are available in very small amounts, with less qualification. Since the issuance of the Metabolites In Safety Testing (MIST) guidance [28], the sourcing and qualification of metabolite reference standards has become increasingly important. More on this topic is presented later. Guidance requires that reference standards come from the USP (United States Pharmacopeia) compendial standards, a reputable commercial source, or are custom synthesized and qualified by an analytical laboratory or other noncommercial establishments [9]. Source and lot number, expiration date, certificates of analyses (if available), and evidence of identity and purity must be furnished.

Biosynthetic or isolated metabolites yield quantities that are insufficient to be as well characterized as synthetic reference materials. One approach isolates microgram levels of individual metabolites for identification using MS and NMR, with quantification using NMR [29,30]. NMR signal response is independent of structure, allowing isolated metabolites to be quantified by comparing their response to that of a known added amount of parent drug. In another approach, radiochemical detection has been used to calibrate stock solutions of metabolites [31]. This approach requires separation of metabolites, as, unlike NMR or MS, detection is nonspecific. Detector response serves to calibrate the concentration of each metabolite by the response to parent drug. The bioanalytical standard curve is prepared by spiking the isolate into the matrix with subsequent dilution. The value of radiochemical detection is that metabolites do not need to be isolated or identified before their quantification. However, radiolabeled drug is needed. For measurements in humans, the metabolite also must be seen in animals or from human *in vitro* experiments. Neither approach requires time-intensive synthesis. However, without synthesis, there is no confirmation of the assigned structure or supply for testing of activity or toxicity. Bioanalytical and biotransformation scientists need to carefully decide how to undertake metabolite quantification to properly integrate approaches with clinical ADME studies.

When measuring endogenous compounds such as biomarkers, the use of a surrogate analyte may be required. Removing endogenous levels by charcoal stripping can be ineffective [32]. Likewise, antibodies may be unavailable for immunoprecipitation. Addition of a stable label analog can serve as a surrogate calibrator. For instance, a tris ^{13}C stable isotope is measured for standard and QC samples. The unlabeled isotope is measured in study samples. To avoid primary kinetic isotope effects, use of a ^{13}C or ^{15}N stable label is preferred. If a ^{2}H stable label is used as a surrogate analyte, its response must be demonstrated to be equivalent to its nondeuterated form.

Another variation of the surrogate analyte approach is when a peptide digestion product is measured as a surrogate for a protein. Protein is indirectly determined by measuring a representative peptide following enzymatic digestion. In one example, matrix metalloproteinase-9 (MMP-9) was determined in mouse serum by capturing on magnetic beads coated with MMP-9 antibody, trypsin digestion, and LC-MS/MS analysis of the surrogate peptide GSPLQGPFLTAR [33]. Assessment of sequence specificity of the selected fragment was aided from database searching. Stable label of the peptide can serve as IS, and pure peptide is used as reference standard. Highly accurate peptide assays can be established. However, consistent capture and conversion of the protein to its surrogate peptide is needed for accurate analysis of protein. Any variation in recovery or conversion of a study sample from the standards will result in errors. Addition of IS following immunoprecipitation and enzyme digestion only serves to compensate for variations in the LC-MS analysis. Similar concerns are present when assaying antibody–drug conjugates (ADCs) for total drug [34]. A more accurate, but less feasible, approach is to incorporate stable label protein before isolation and digestion.

17.3.2 Quality Controls

QC samples should be positioned evenly through the batch, reflecting the concentration of study samples. When study data fall over a small percentage of the calibration curve, none of the QC concentrations may be near the unknown concentrations, limiting the

monitoring power of the QC samples. Adjustment of either the assay range or QC number and placement is expected.

QC samples are measured at a minimum of three concentrations in duplicate (or at least 5% of the unknown samples) are required. Spacing is within 3× LLOQ, one near the center and one near the upper boundary (top quartile). Grouping "near the center" or "midrange of the calibration curve" has afforded some latitude in mid-QC placement. Some have moved from its linear (50% of ULOQ) placement closer to a logarithmic (geometric mean) placement. This is consistent with a desire to more accurately capture the elimination phase. Depending on interpretation, there are significant differences. Table 17.1 illustrates the disproportionate gaps between QC samples when a linear placement is used instead of a geometric mean. When calculating the mid-QC, adjustment of its final concentration should be done to allow for easy preparation of this QC. Another option is to include two mid-QC samples. This has the added value of providing better coverage when study samples differ from their expected concentrations. Wider-range assays (≥1000-fold) should include the addition of more QC levels to better represent study data.

During sample analysis, each batch should contain at least six (minimum 5% of total) QC samples. QC samples at a minimum of three concentrations in duplicate are added to each plate. At least two-thirds overall and 50% at each level must be within 15% of their nominal value. It is advantageous to avoid odd replicates, as at least two of three QC samples would be required to meet the 50% rule.

The analyst should make every effect to minimize the amount of organic spiking solution and to match QC and standard preparation. Study samples contain no organic solution. Even a few percent of an organic solution can denature proteins, causing a loss in enzyme activity or protein binding. The result is an inaccurate assessment of stability. Deactivation yields a false stability not seen with incurred samples. Apparent instability due to analyte entrapment within precipitating plasma proteins or instability when analyte is released from protecting proteins can also occur. Preparation by serial dilution in matrix is one effective means to reduce the organic composition in QC samples.

When performing an assay validation, more replicates and additional QC samples are required to assess accuracy and precision at the LLOQ and to test dilutional linearity.

TABLE 17.1 Dynamic Range and its Associated Gap Between QC Samples Using Either Linear (Top) or Geometric Mean (Bottom) QC Placement of Middle QC

Range	Low	Mid	High	Fold Gap (Low–Mid)	Fold Gap (Mid–High)
Linear Mid-QC					
100	3	20	80	5.7	3.0
250	3	50	200	15.7	3.0
500	3	100	400	32.3	3.0
1000	3	200	800	65.7	3.0
GM Mid-QC[a]					
100	3	15.5	80	4.2	4.2
250	3	24.5	200	7.2	7.2
500	3	34.6	400	10.5	10.5
1000	3	49.0	800	15.3	15.3

Abbreviation: GM, geometric mean.
[a]Square root (low QC × high QC).

A minimum of five determinations at each QC concentration are required. In practice, many laboratories use replicates of six to better align with acceptance criteria. Additional QC samples include the LLOQ QC whose intraday and interday precision and accuracy needs to be within 20%. Dilution QC samples should test the maximum range over which the dilution is anticipated. If proven during validation, dilution QC samples do not need to be run in study sample analysis. This approach tests only dilutional linearity of the assay. It does not provide routine monitoring of the analyst's ability to execute dilutions nor does it test the ability to dilute matrix from different subjects (parallelism). For that reason, many include dilution QC samples within runs to support values reported for any diluted samples. However, neither LLOQ nor dilution QC samples need to be included within regular analytical runs.

While up to 25% of standards not meeting acceptance criteria are deactivated, QC samples should never be removed unless there was an assignable analytical cause. Measurement of the intrarun and interrun assay precision and accuracy during both validation and study sample analysis requires the entire population of QC data. Outlier tests such as Dixon's or Grubbs' [35] test may be applied, but they should be judiciously used.

QC samples are considered as a surrogate for study samples in which the true (nominal) concentration is known. Testing of incurred (study) samples reflects the observation that QC samples do not always behave the same as study samples. Some causes for this observation are discussed later. While replicate analysis of incurred samples tests interday precision, QC samples generally provide the only assessment of accuracy. Additional means to derive a true result from an incurred sample include measurements at multiple dilutions (tests parallelism) and the method of standard addition. Both require multiple aliquots of sample, are time intensive, and would need validation. For these reasons, they are not generally used. Validation establishes precision using QC samples selected by the bioanalyst, whereas ISR tests interday assay variability from the study population.

17.4 MATRIX EFFECTS

Nonselective detection of analyte signal, resulting in the apparent elevation of its true concentration, is referred to as *background* or *chemical noise*. Components within the chemical matrix can also either reduce or elevate the apparent analyte concentration by affecting detector response. The matrix effect is therefore the suppression or enhancement of ionization of analytes by matrix components in the biological samples [36]. A quantitative measure is the matrix factor (MF), which is the ratio of the analyte peak response in the presence of matrix ions to its response in the absence of matrix ions (Table 17.2). Spiking analyte into extracted blank matrix and comparing its response to an equivalent amount of pure compound is the best means to quantitatively assess this parameter [37]. The MFs of both the analyte and ISs should be determined. The relative MF is the ratio of the two and should be near 1.0.

Matrix effects are generally due to components within the sample, which are at high abundance, and yield high response. Salts or urinary acids are abundant and ionize readily in ESI, affecting analyte response in urine assays. Urinary salts or acids can also entrap drug during precipitation, so making the proper assessment is important. The high abundance, good surface properties, and ionic nature of phospholipids make them

TABLE 17.2 Determining Matrix Factor and Recovery (Extraction Efficiency)

Analyte	Internal Standard
Preextraction spike	Preextraction spike
Postextraction spike to blank	Postextraction spike to blank
Mobile phase spike	Mobile phase spike

Recovery (analyte or IS) = (preextraction spike/postextraction spike) × 100%
Matrix factor (analyte or IS) = (postextraction spike/mobile phase spike)

a concern in plasma assays [38,39]. Bile acids, present at millimolar concentrations in bile, can impact analyte ionization. The levels of bile acids in urine or plasma are far less and therefore not generally a problem in these matrices. Each matrix and each ionization method has its unique challenges, requiring sufficient extraction or chromatographic separation before ionization.

Ionization that occurs concurrently with desorption from a condensed phase is generally more prone to matrix effects [40]. ESI has shown greater matrix effects than atmospheric pressure chemical ionization (APCI) or atmospheric pressure photoionization (APPI). The requirement to volatilize analyte before either ionization within a corona discharge (APCI) or photoionization (APPI) source effectively removes many matrix components that impact analyte ionization [41,42]. Phospholipids are not volatile and therefore do not present the same problems in APCI or APPI as they do in ESI. However, many analytes are not volatile and require ESI with proper sample processing before ionization.

Indirect detection of matrix effects by postcolumn infusion is relatively insensitive and does not identify the offending component within the matrix [36]. Having a direct and simple means to identify exact phospholipids in plasma greatly assists their monitoring and removal [43]. Positive precursor ion scan of m/z 184, positive neutral loss (NL) scan of 141 Da, and negative precursor ion scan of m/z 153 are the three widely used means. When a total assessment of phosphatidylcholines is desired, monitoring the m/z 184 → 184 transition under high source energy is most commonly used [39].

ISs in MS-based assays can minimize the impact of matrix effects by normalizing response ratios. Matrix effects observed for stable isotope labeled IS are similar to those observed for the analyte, particularly for ^{13}C or ^{15}N and less so for ^{2}H stable label analogs. From a practical perspective, deuterated ISs are generally more readily available and some can be prepared using single step alkylation or hydrogenation reactions. Deuterated stable labels can exhibit primary kinetic isotope effects not seen with ^{13}C or ^{15}N stable labels. They are generally acceptable, provided the label is not scrambled during MS analysis and is metabolically stable within the matrix. The introduction of higher resolution separations (e.g., UPLC (ultraperformance liquid chromatography)) and more extensive substitution can result in uncompensated matrix effects if analyte and IS elute at different times. For these reasons, some have suggested coelution is always preferred, including when using a structural analog [44].

Matrix effects are less a problem when the matrix of the study population matches that used to prepare standard and QC samples. If the analyte response is affected in study samples, it should be similarly affected in standard and QC samples. Detection sensitivity can be affected but accurate quantification would not be. That premise is

generally true for animal studies, in which strains of similar genetic and environmental background are tested.

Matrix effect measures response from a sample of blank matrix to which an equivalent amount of analyte was added after extraction (postextraction spike) to response from a sample that contains the same amount in mobile phase (no matrix). Having a highly variable MF in individual subjects would be a cause for the lack of reproducibility of analysis. It is therefore required that the absolute MF and IS-normalized MF for six individual lots of the matrix be tested and that the variability in MFs, as measured by the coefficient of variation (CV), be within 15%. However, when using stable isotope IS, it is not necessary to determine the MF in six different lots.

To test assay specificity, Brazilian requirements define that samples of the biological matrix must be analyzed from at least one lipemic and one hemolyzed sample [16]. Blank samples must be tested using the established procedure and results compared with those obtained from the aqueous solution of the analyte, in concentration close to the LLOQ. A similar assessment within the EMEA guidance requires measuring MF in QC samples prepared from both lipemic and hemolyzed plasma [15]. This assessment will test the potential for phospholipid suppression at its extreme.

17.5 RECOVERY

The introduction of requirements to test sample diversity during method development and validation is good assurance that methods either will be entirely insensitive to population variability or will define their failure points. Brazilian guidance also requires that at least one hemolyzed sample be tested for specificity in a manner similar to that noted above [16]. While spectral interference from hemolyzed red blood cells (RBCs) is common in colorimetric assays used in clinical chemistry, the specificity of LC-MS assays makes them less prone to this problem. There are cases, however, when an extensively hemolyzed sample measured using LC-MS will be affected. The cause may include instability following the release of RBC enzymes, lower recovery from cell debris, or matrix effects from RBC components. The EMEA guidance requires measuring MF in at least one low QC sample prepared from hemolyzed plasma [15]. Testing extensively hemolyzed plasma assures one that lysed blood cells or their intracellular contents do not impact analyte stability, recovery, or detection. If the assay is sensitive to hemolysis, knowing its limits would allow one to make assessments of whether data should be reported [45,46]. Since most samples are not measured for hemolysis, attention to sample conditions and good documentation is needed.

Having an assay that meets global (FDA/US, EMEA/Europe, ANVISA/Brazil, TGA/Australia, SFDA/China, and MHLW/Japan) requirements is essential to support the registration of new drug candidates in all countries. Some of the less common requirements are associated with how matrix effect and recovery are assessed. There is an extensive effort among the bioanalytical communities to ensure that common standards are adopted. Included in this effort are the Global Bioanalysis Consortium, European Bioanalysis Forum, and Global Contract Research Organization Council for Bioanalysis. Large challenges lie ahead to achieve the right balance of consensus and the appropriate level of detail in the upcoming EMEA and FDA Guidance.

To accurately assess the recovery (extraction efficiency) of analyte or IS from the sample matrix using MS, matrix effects must be eliminated. This is achieved by

comparing response from an extracted sample to a sample of blank matrix to which an equivalent amount of analyte was added after extraction (Table 17.2). Use of a postextracted spiked sample corrects for matrix effects. Low, mid, and high QC levels need to be tested for recovery. Apart from the criteria that recovery should be consistent (independent of concentration), precise and reproducible for both the analyte and IS, there is no required minimum. Higher recovery is better, as even small changes in a low recovery assay can have great impact. Some laboratories do require a minimum by standard operating procedure (SOP).

The use of a postextraction spike works well for off-line extraction, but what about on-line methods such as turbulent flow [47]? When matrix is directly injected onto an on-line extraction with subsequent LC-MS analysis, an isolated extract is not obtained. In this case, the extraction recovery is presumed to be quantitative. To comply with the draft EMEA guidance, interbatch variability in response should be assessed by analyzing at least six batches of matrix in triplicate and achieving an overall CV within 15% [15].

Provided a close structural analog has been chosen as the IS, comparable recovery should be seen. Indeed, this should be one of the many defining criteria used in selecting an appropriate IS. Whenever possible, a stable label is preferred. Even still, it is important to ensure that the IS is properly equilibrated with the analyte in the matrix before beginning extraction. High affinity binding to matrix components can be difficult to disrupt. For instance, 2 mM of a competitive displacer was incubated for 2 h at 37°C to fully displace nanomolar levels of roxifiban from platelets [48]. Even the use of a stable label will not compensate for a procedure that improperly equilibrates IS before extraction.

Achieving consistent recovery from plasma or serum is less problematic than when one is working with blood, cells, tissues, or hard matrices such as bone. Partitioning between blood cells and plasma is important to measure for two reasons. First, it may determine the choice of matrix, blood, instead of plasma. Second, time-dependent partitioning is possible, so differences in the harvesting of plasma will impact PK. Similarly, slow release of analyte from tissue homogenates can complicate analyses. Proper introduction and equilibration of IS before extraction is needed to overcome differential recovery [49]. Recovery from study samples that have undergone different periods of equilibration should be tested. Slow equilibration of IS with analyte in the matrix may require a longer equilibration time, more vigorous mixing, or the addition of acid or a competitive displacer to overcome high affinity binding. Matrices such as bone require complete digestion before analysis [23]. Matrices that are more complex and difficult to homogenize will always present challenges to prepare standard and QC samples that mimic study samples. For complex cases, dosing radiolabeled drug and assessing recovery from study samples is an alternative approach.

Through-LC recovery, which for absorptive compounds can be greatly reduced at lower concentrations, is generally not tested or known. When low level adsorption is seen, the standard curve may have decreased sensitivity at lower levels than that seen from upper standards. Using high levels of an appropriate IS may compensate by masking active surfaces and exhibiting similar recovery at lower levels. It is more appropriate to select a better stationary or mobile phase that overcomes these low level losses. Peak tailing and varying background are often associated with reduced through-LC recovery at low concentrations. Both symptoms illustrate a propensity for high affinity binding to the injector or stationary phase. Including additives to the

mobile phase can overcome losses, but it must be done with care to avoid affecting MS response. The lack of understanding and correcting through-LC recovery problems can result in sensitivity changes during the run, poor apparent detection sensitivity, and variability in the chromatographic baseline. It is one reason why an extensive number of conditioning samples may be needed in an assay.

17.6 CARRYOVER

Carryover in an HPLC system is contamination by the preceding sample. It is generally predictable and quantifiable by examination of the method, allowing an accurate assessment of its impact. Method development should define the individual sources (e.g., liquid handler vs HPLC) so that its causes can be eliminated or reduced. Method validation should accurately measure carryover to define its expected contribution when samples are assayed. Relative carryover is measured by injecting the following sequence, blank-ULOQ-blank. An acceptable result is when the difference of the blanks is $\leq$20% LLOQ. If carryover is more than 20% LLOQ, its acceptance should be justified and appropriate precautions used to minimize its impact.

A direct, yet inconvenient correction is to restrict the dynamic range of the assay. Precautions include the proper placement of samples to minimize changes in concentration and protection of samples with washout blanks. Even still, the influence of carryover is unknown when the preceding sample is above the limit of quantification (ALQ). It is not until the ALQ sample is reassayed that an assessment can be made. Concentration-dependent carryover can complicate this assessment. Saturation of nonspecific binding sites causes concentration-dependent carryover and is most evident when a larger relative carryover is seen on subsequent injections (delayed washout). One can implement procedures that recognize high carryover within a run and *a priori* rules to reject individual samples or batches [50]. If not reduced and uncontrolled, one should be prepared to fail individual results or runs, intersperse blank samples, and state its potential impact on study results.

For these reasons, there is considerable commercial effort to reduce the absorptive nature of materials used in HPLC systems, particularly within autosamplers. Improvements include using more inert materials, avoiding contact with the syringe needles, and improving needle or valve washing. The present carryover criteria have all but eliminated the direct injection of matrix in bioanalysis, as this approach often results in the buildup of absorptive components [47].

Carryover forced a revision of earlier guidance that recommended randomizing samples. Randomization could result in low level (24 h) samples immediately following a high level (2 min IV) sample. In this example, the preferred sequence to minimize carryover would be to start assaying the latest time point samples and move to earlier ones.

17.7 DILUTIONS

Guidance does not mandate within study dilution QC samples to be used, provided dilutions are conducted with like matrix and tested within validation. The extent to which study samples are diluted must be tested during validation. Although allowed,

this approach does not check that dilutions were properly done during study sample analysis. A preferable approach is to include dilution QC samples whenever required by study samples.

When drug levels exceed the range of the highest dilution, additional dilution factors can be tested during sample analysis. If dilution is performed with an unlike matrix, QC samples must be diluted in the same manner as the study samples and analyzed with the diluted samples. All diluted QC samples should be created within the assay calibration range.

A few laboratories employ a procedure that measures the same sample at multiple dilutions within the same run. This procedure is more common when drug levels are unknown, high dilutions are needed, and/or curve ranges are limited. While this approach can provide more confidence in assay results by replicate demonstration of parallelism, *a priori* acceptance and reporting rules must be established.

17.8 STABILITY

Procedures must evaluate the stability of analytes during sample collection and handling. Conditions used in stability experiments should reflect situations encountered during actual sample handling, including the collection of plasma from blood. A common test is to process plasma or serum during blood collection as shown below:

- spike fresh blood with analyte(s) at low and high QC concentrations
- incubate at ice, room, and body temperatures (or warm blood to 37°C before room or ice processing)
- harvest plasma after 0, 30, 60, and 120 min of storage using a nonrefrigerated and/or refrigerated centrifuge
- transfer plasma to fresh tubes, and store frozen until analyzed together in a single set.

When analyzed, the compound is considered unstable when the measured concentration is less than 85.0% of the initial concentration. If performed at 37°C, this procedure can also be used to assess blood-to-plasma partitioning, as T0 plasma concentrations different from the nominal blood concentration reflect partitioning. As noted previously, knowing the blood-to-plasma ratio can define the assay matrix. For drugs that are highly partitioned into blood, analyte should be measured in blood and not plasma. Highly unstable compounds or those showing time-dependent blood-to-plasma partitioning will require special attention. Testing blood stability in all species is preferred since blood-to-plasma partitioning can be time and species dependent and enzyme levels vary across species. Blood should be freshly collected to avoid loss of enzyme activity.

Esterase activity in rodents is elevated but reduced in higher animals and man. For prodrug and acyl glucuronide assays, conditions that stabilize mouse or rat blood are generally sufficient for human, dog, or monkey studies. However, it is common to have difficulties if the procedure used to establish stability in humans is assumed to be sufficient for rodents. Likewise, oxidases and other enzymes can show similar species specificity. Thiols that form mixed disulfides can be challenging and require conditions to stabilize their breakdown back to parent drug. Owing to interspecies differences in

enzyme activity, it is preferred to determine stability in all test species at the beginning of method development.

Sensitivity to light is often known or easily tested, and hence, yellow lighting for sample processing and amber containers for storage are commonly required. Chemical and light stability assessments have generally been performed on new drug candidates, so this information should be available.

The effect of different anticoagulants on stability should also be considered. EDTA, a common anticoagulant, can be an effective phosphatase inhibitor. Phosphate ester prodrugs collected in EDTA plasma are stabilized. Heparin does not afford the same protection. Likewise, a drug that is a strong chelator may yield different apparent stability in EDTA than when using a nonchelating anticoagulant. Heparin can afford stability by binding drug while EDTA cannot. The stability in different anticoagulants should not be considered equivalent and must always be tested.

When measuring metabolites, it is essential to use an assay that is specific for the metabolite and to employ proper stabilization of any phase II conjugates or interconverting forms (e.g., lactone → acid). Phase II conjugates are common. Fortunately, their physiochemical properties often afford a relatively simple separation. The same is not true for phase I metabolites. It is common to have multiple hydroxylated forms, of which either one or many may maintain activity and be of interest for routine measurement. It is critical that all others (and their phase II conjugates) are separated from the metabolites being measured. During method development, it is helpful to obtain a source of known metabolites and spike them into the matrix. Preliminary stability tests using changes from initial response can guide decisions on what stabilization is needed.

The use of selective inhibitors is preferred since these can be added in limited quantities and can result in minimum change to the matrix. This assumes that the general enzymatic process is understood. For esters and lactones, esterase inhibitors such as diisopropyl fluorophosphates (DFP), phenylmethylsulfonylfluoride (PMSF), and 2-thenoyltrifluoroacetone (TTFA) can be effective. For phosphate esters, EDTA and Halts phosphatase inhibitor are commonly used. For oxidases, antioxidants such as ascorbic acid can be effective. Amides may have sensitivity to esterases, but they can also be hydrolyzed by fatty acid hydrolase or proteases. For peptides or proteins, a general protease inhibitor such as aprotinin is often needed. Since an understanding of its metabolic fate may still be evolving for new chemical entities, some screening of different classes of stabilizers may be needed [51].

Many peptides and proteins show high nonspecific binding, requiring the use of special sample collection tubes such as the Nunc™ immunoassay tube. Adsorption losses may be misinterpreted as instability. To minimize losses, testing of containers should be considered. Adsorption losses are common in matrices such as urine and CSF and may be overcome using surfactants (CHAPS, Tween-20, Triton X-100), protein (BSA), or EDTA to minimize surface contact or using additives (organic, acid, or base) to enhance solubility. If samples were collected before a procedure is established, it may be possible to recover the absorbed compound from the collection tube. In other cases, drug is irreversibly bound and cannot be recovered. For urine, generally smaller aliquots, rather than the large primary collection vesicle, are stored. Laboratory tests should determine whether or not additives are needed at the collection site. It is a mistake to not test drug adsorption onto urine or CSF containers before a study is started. The procedure should account for any dilution on addition of a stabilizing

solution. Provided they can be properly aliquoted, concentrated solutions are preferred, as the dilution is limited. If a fixed amount is added to the container, one should test extremes in urine volumes. Although it is better to adjust the aliquot to the weight or volume of the urine collection, this is often impractical in the clinic [52].

Use of a low pH (3–4) acid or buffer (formic and citrate) affords optimum chemical stabilization of esters or lactones. Lowering the pH is also important to avoid migration or anomerization of acyl glucuronides [53]. Direct assays for acyl glucuronides must ensure specificity to the 1-β isomer, as one cannot assume that other migrated isomers yield the same MS response. It is therefore important that the 1-β anomer be resolved from all other forms [54]. One should always consider reducing the pH when collecting samples of an acid-containing drug to avoid the breakdown of any potential acyl glucuronides. However, there can be times when the drug itself is less stable under these conditions, so some preliminary stability assessment is warranted before defining sample collection conditions.

When labile ester-containing drugs are measured, both an esterase inhibitor and low pH may be needed. Using a high concentration of acid or buffer alone to denature all enzymes should be avoided. What stabilizes drug and metabolite may destabilize the matrix and complicate the assay. Stabilizers added at high concentration can result in difficulties because of inhomogeneous sampling and poor recovery from a viscous matrix. Hemolysis and gelling are far more common when adding stabilizers. Difficulties in sample collection at clinics may not be predicted from laboratory experimentation. Stabilizers may also cause interference, either directly as chemical interference or indirectly by suppressing ionization (matrix effects).

When adding stabilizers, it is important to correct for any dilution. Stabilizer added to blood within the Vacutainer™ will require an assessment of its partitioning into both blood and plasma. This is generally assumed to be, when corrected for hematocrit, equal. When preparing tubes, it is essential that vacuum is maintained and a regular assessment of the effectiveness of the stabilizer is performed. If plasma or serum harvesting can be done without the need to add stabilizer to blood, then a simpler dilution is possible. In practice, the use of a commercial tube that contains a solid stabilizer (e.g., sodium fluoride) is preferred, as this avoids dilution and preparation issues. While sodium fluoride may afford adequate esterase protection for some drugs in certain species, it may be inadequate for others. This is commonly the case in rodents, so special collection procedures may be needed for rodent studies, whereas a commercially available tube containing sodium fluoride may work for clinical studies.

Once proper means to measure drug levels and stabilize samples have been achieved in method development, validation can proceed. Validation will include appropriate assessments to determine stability. No sample can be reported without ensuring that stability determinations covered the period of sample storage and processing. Stressed (minimum low and high) QC levels must be within 15% of their nominal value. Assessments in the unaltered matrix of qualifying QC samples include

- long-term stability under storage conditions
- freeze–thaw cycles
- bench-top (room temperature and/or ice processing)
- processed sample or extract stability
- reinjection integrity.

Details on the stability procedures are given within the references and are not repeated here. Stability assessments can be an ongoing process as the studies progress. Most laboratories will require a minimum of three freeze–thaw cycles, six or more hours of bench-top stability and extract stability, and two weeks of long-term stability within the validation. If the sample extract is to be reanalyzed, successful reinjection of an acceptable run is required (reinjection integrity). This is different from an extract stability assessment, which is demonstrated using a fresh standard curve.

The stability of reference standards and IS stock solutions must also be demonstrated under storage and bench-top conditions. Secondary stock solutions not immediately used must also be tested. An apparent degradation of more than 10% is unacceptable. Mixtures of stock solutions can be more problematic, as common ion effects may limit solubility. Frozen stock solutions are to be avoided. Testing of different solvents and containers during method development is good insurance that stability assessments in validation will be successful.

17.9 SPECIFICITY AND SELECTIVITY

FDA guidance states that the specificity of the assay methodology should be established using a minimum of six independent sources of the same matrix [9]. This assessment is made by ensuring that background response from six lots of blank matrix is $\leq$20% of the response from the LLOQ sample and $\leq$5% of the IS. Unless a nonselective channel is monitored or analysis at low levels is performed, LC-MS-MS chromatograms generally show few interfering peaks.

FDA guidance recognizes that for MS-based methods, testing six independent matrices for interference may not be definitive [9]. In the case of LC-MS assays, matrix effects should be investigated to ensure that precision, selectivity, and sensitivity will not be compromised. Selectivity is defined within the guidance as "the ability of the bioanalytical method to measure and differentiate the analytes in the presence of components that may be expected to be present." Guidance also requires that "each blank sample should be tested for interference, and selectivity should be ensured at the lower limit of quantification (LLOQ)." The most definitive test is to measure the LLOQ concentration from six different lots of matrix to $\pm$20% acceptance.

Laboratories need to decide whether a matrix lot that fails can be reanalyzed and how many additional lots should be tested. A reasonable rule would ensure that if more than 10% of tested lots fail, the analyst should consider redeveloping the method. Owing to the specific detection employed in MS, fewer lots of matrix are screened for interference and assessed for selectivity than when using less specific assays such as immunoassay. Accurate quantification may be affected by causes other than nonspecific detection, so measuring QC concentrations in at least six different lots of matrix is required to support assay selectivity.

An assay should be tested whenever clinical studies are performed, when there is the potential for concomitant medication, or when it is used in DDI studies. Over-the-counter (OTC) interference assessments include a mixture of commonly used medications at an appropriately high concentration. For prescribed medications or in DDI studies, tests of the other components are typically done at the anticipated C_{max} or at a higher value. The potentially interfering drugs and their significant metabolites should be added to the blank matrix and the assay tested to ensure that there is $\leq$20%

of LLOQ response and ≤5% of IS response. This assessment tests specificity or the direct interference as noted from background response. Infrequently, comedications may cause an indirect effect if they are abundant, may coelute, and are preferentially ionized. Potential interferences in LC-MS are often predictable based on their molecular weight.

Guidance also states that "potential interfering substances in a biological matrix include endogenous matrix components, metabolites, decomposition products, and in the actual study, concomitant medication and other exogenous xenobiotics. If the method is intended to quantify more than one analyte, each analyte should be tested." Finding a source for these potential interferences can be a challenge. However, the analyst is expected to ensure that every attempt was made to secure and test them. Compounds are added at physiological concentrations to the matrix, fortified with analyte(s), and analyzed. Normal acceptance rules (four of six QC samples must be within 15% or 20% at LLOQ) apply.

The suitability of an assay to measure drug levels in special populations (age/gender, renal/hepatic impaired) should be considered before starting sample analysis. Differences in the clinical chemistry of special populations can impact assay performance. Elevated phospholipids in lipemic populations can cause matrix suppression not apparent in standards prepared from normal plasma. Hemolysis may be more prevalent or extensive, particularly when special processing is required or remote clinics are used. Accumulation of metabolites in slower clearing populations may result in a metabolite interference not previously seen in normal or healthy populations. It is therefore important to have an understanding of the anticipated metabolism before making this assessment. Assay levels may also need to be adjusted to better cover these populations.

Since biotransformation generally involves a change in the molecular weight of the drug, interference from metabolites when measuring parent drug levels is less frequent than when measuring metabolites. Phase II conjugates present concern when coeluting ether glucuronide fragments within the ion source or when acyl glucuronides degrade in-process back to parent drug. Either process results in overestimating parent drug exposure. More common is to have many isomeric or isobaric metabolites, all of which should be resolved from the metabolite being measured. For isomeric metabolites, this generally requires chromatographic separation [55]. For isobaric metabolites, a specific multiple reaction monitoring (MRM) fragmentation is more likely to be achieved. Analysts should take caution in assessing the specificity of a method to measure metabolites. Access to samples from a previous or pilot study helps and should always be considered as a precautionary step. Alternatively, metabolites from *in vitro* sources or animal *in vivo* studies can be spiked into the matrix to assess their potential for interference. Testing from a prior or pilot study is preferred since it affords the best assessment of analytical relevance. The spiking approach, however, should be considered for first-in-human studies.

17.10 CONTAMINATION

In contrast to carryover, contamination is unpredictable and may occur during any part of the study (dosing, sample collection, or analysis). We focus first on sample analysis and later illustrate means to assess study contamination. Contamination may occur in a

limited number of study, standard, or QC samples and to a varying extent. Therefore, assessing its impact can be difficult.

Contamination has become more problematic as instrument sensitivity has increased and samples are processed in closer proximity. For example, a tube-based, HPLC-UV assay that measures microgram per milliliter levels does not present the contamination challenge of a 384-well SPE (solid-phase extraction) LC-MS/MS assay that measures picogram per milliliter levels. Well-to-well or cross-contamination during sample preparation can be reduced by proper programming of the liquid handler method to avoid dripping or tip-to-tip contamination. Disposable pipette tips, although more expensive than fixed tips, can overcome contamination. If fixed tips are needed, appropriate wash steps should be performed. Placement can be considered to minimize its impact; however, isolating high level samples from neighboring wells is impractical.

Other general considerations include separating weighing rooms, where high level stock solutions are prepared, from the LC-MS laboratory. Likewise, bench areas should be covered with disposable paper and instrumentation regularly cleaned to avoid system contamination. Reagent contamination (matrix, solutions), at very low levels, can be difficult to recognize as analyte contamination versus chemical background. An easy test includes adding more MRM transitions so that several profiles are required to have the same relative abundance as analyte. By increasing assay specificity, the analyst can be assured that the problem is contamination and not simply chemical background. For chemical background, an additional transition may serve as the detection channel in a revised method.

If contamination during extraction is suspected, one may use cross-contamination markers to indicate well-to-well contamination during sample processing. Compounds are added in a checkerboard pattern and analyzed to note adjacent well contamination [56]. When used, the analyst must prove that the markers do not influence the assay. The impact must be judged by the bioanalyst using rules defined within an SOP. Options include rejecting the sample(s) or batch, raising the LLOQ, or tolerating the bias (contamination $<$ experimental error).

The bioanalyst has the responsibility to resolve any questions regarding sample analysis. The study director has the responsibility to determine the overall cause(s) for contamination and assess its impact. When needed, the analytical method can also serve to trace contamination within the animal facility or from other sources. Confirmation by reanalysis can generally resolve analytical questions, provided the original sample was not contaminated at first sampling. Given that dosing often occurs at concentrations up to 1,000,000-fold greater than bioanalysis, even a small fraction of dose transferred into or about the tube can find its way inside the sample. For this reason, many laboratories take second samples to avoid random errors associated with sampling contamination.

Following issuance of the EMEA Guideline on Control Animals, more attention was paid to resolving contamination issues [21]. For all pivotal studies that include a toxicokinetic evaluation, control samples should be analyzed. In nonrodent studies, full profiles are required. In rodent studies, sufficient samples about the T_{max} are needed. When significant contamination is at a level that can impact the validity of the study, the sources of contamination should be investigated. This includes independent analytical and in-life investigations, which are reviewed and reconciled by the study director. It is the responsibility of the study director to assess the significance of impact (greater than no adverse effect level) on the study results. Options to remediate the problem in

future studies should be assessed and implemented. Findings should be stated in the written summaries and highlighted in the nonclinical overview.

When assessing contamination in control animals, the assay range should be considered. To minimize spurious signals being detected as contamination, the LLOQ should be set at an appropriate level. One approach for trough sampling targets an LLOQ that is a percentage of the concentration expected to occur at 24 h after the lowest dose. Interanimal variability must be considered to ensure that the lowest exposed animal within the low dose group can still be measured.

The objective is to determine whether control animals were exposed to drug. Therefore, control samples can be run on a separate plate to minimize the potential of analytical contamination. Contamination sentinels, or additional blanks, can also be used to indicate analytical contamination. An appropriate acceptance rule would require that all contamination blanks included in each analytical run do not yield measurable values. If blanks are below the limit of quantitation (BLQ) and toxicology control samples show quantifiable exposures, control samples can be reassayed in replicate to both confirm analysis and define a reportable value. Confirmation of animal exposure can be accomplished by monitoring for metabolites. For metabolites assayed with parent drug, their ratio should be different from standards or change over time to ensure the lack of sample contamination during analysis.

17.11 SYSTEM SUITABILITY AND RESPONSE CHANGES

Only qualified and properly maintained instruments should be used for drug analysis. To qualify instruments at time of use, a system suitability test may be used. System suitability samples test that the instrument is sufficiently sensitive, selective, and reproducible for the current analytical run. This generally includes replicate testing of the analyte for sensitivity and precision near the LLOQ before starting an analytical run. Similar tests following the run may be used to confirm instrument stability. A mixture of components can test multiple analytes for assay specificity, such as the resolution of critical pairs in a separation. System suitability tests do not replace the required run acceptance criteria. They can be effective in helping to decide whether sample extract can be simply reinjected or a new extraction and reanalysis are needed.

A means to judge individual results, based on the response of ISs, is often dependent on the assay. However, most laboratories have developed general rules that allow *a priori* decisions to reject sample results that show aberrant IS response. The rationale is that any significant change in observed response can reflect an unknown error in sample processing (double spike of IS solution) or an assay change (recovery or matrix effect, injected volume, or instrument response). While an assay may be rugged enough to compensate for an error such as a reduced injection volume, it is sometimes difficult to differentiate an apparent response change from a more fundamental problem. The criterion that only a significant ($>50\%$) change in an individual IS response from the mean IS response is generally used to reject sample results.

Figure 17.1 illustrates an example in which conditioning of the instrument was insufficient before starting an analytical run. This example also illustrates how an individual subject profile may be different from others within the study or QC samples. Today, most laboratories use this type of assessment to help troubleshoot assay runs and make proper decisions on whether samples can be reinjected, reassayed, or not

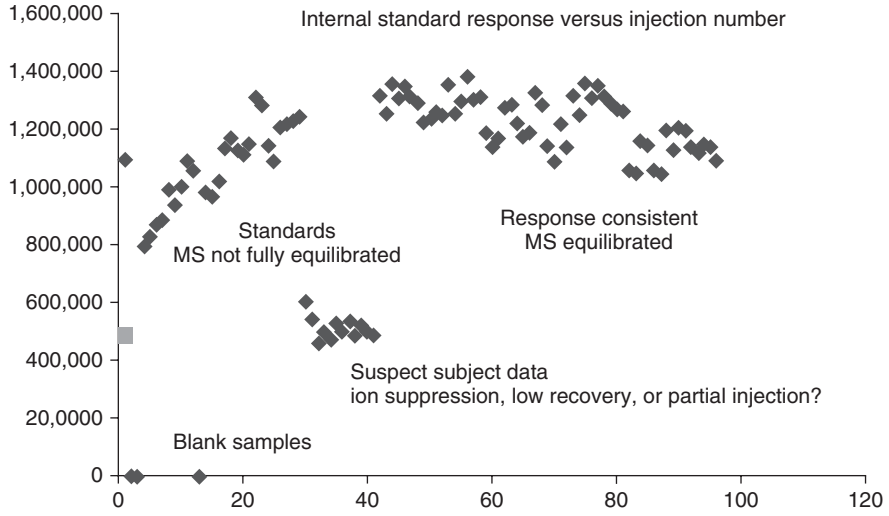

Figure 17.1 Plot of internal standard response versus injection number illustrates the need for proper conditioning (equilibration) of the LC-MS instrument before starting an analytical run. It also allows assessment of how individual subject data should be treated and when MS sensitivity changes.

reported. Some laboratories include other within-run monitoring such as phospholipid assessments to determine whether ion suppression occurred.

Figure 17.2 illustrates an example where a less potent derivatization reagent was uncovered by monitoring changes in d_3-Δ9-tetrahydrocannabinol IS response. Aged QC samples showed a pronounced difference from fresh standard or QC samples, suggesting either poor IS recovery or less extensive derivatization. When another bottle of derivatization reagent was used with more vigorous vortexing, there was no difference in response between fresh and aged QC samples. The fortunate use of a stable label IS compensated for these changes, as indicated from standard and QC samples. Use of a stable label rather than a structural analog as IS is always preferable, particularly for more complex extractions and derivatizations. Provided the IS is well equilibrated with analyte in the matrix, a stable label will compensate for any subsequent changes. When using stable label ISs, having additional rejection criteria are less critical.

17.12 SAMPLE REANALYSIS

Samples may be reanalyzed for an assignable cause such as a technical error, poor chromatography, unacceptable IS response, or evidence of carryover. The original result is discarded, and the sample is reanalyzed in singlet. The acceptable reanalysis value is reported.

Samples may also be reanalyzed for anomalous results or for PK reasons. These analyses are performed in a manner similar to replicate confirmations of contamination. Typically, confirmatory reanalysis is performed in duplicate. Provided the two replicates agree, a comparison is made between their mean and the original result. Alternatively, the median of the three may be reported. If any one of the three results is below the

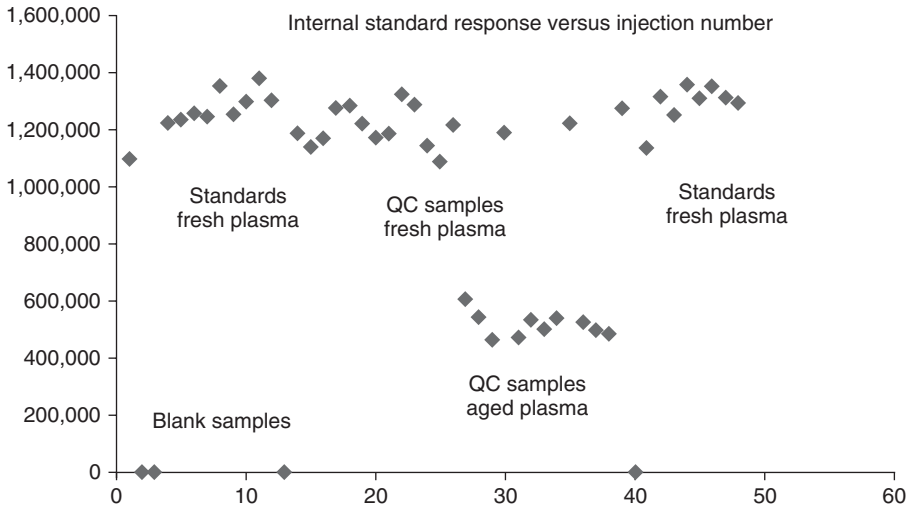

Figure 17.2 Plot of d_3-Δ9-tetrahydrocannabinol IS response versus injection number for the analysis of Δ9-tetrahydrocannabinol in human plasma illustrates how the extent of derivatization was reduced in aged plasma when using a less potent solution. Fresh reagent and more vigorous vortexing returned IS response of aged to normal plasma (not shown).

LLOQ, the less of the two remaining results is usually reported. When two or more are below the LLOQ, the result is typically BLQ. If only sufficient volume remains for a single determination, a comparison is made between the two results. Either of the three events may occur: the original result is confirmed and reported (<15%), a mean value is reported (15–30%), or no value is reported (>30%). There are several acceptable means to judge replicate repeat analysis. A decision tree, established within an SOP, should define the procedure.

There are a number of situations where the performance of QC samples may not adequately mimic that of study samples from dosed subjects (incurred samples). Reasons for the difference can be considerable, including metabolites converting to parent drug, protein-binding differences in patient samples, recovery issues, sample inhomogeneity, matrix effects and interference from formulation, or drug-modulated differences in clinical chemistry. These factors can affect the reproducibility and accuracy of the concentration determined in incurred samples. Since accuracy is difficult to determine in a study sample, replicate analysis (interassay precision) is used as a surrogate for accuracy. Irreproducibility, therefore, exposes inaccuracies, which warrant further investigation. While the causes for irreproducibility are often characterized and minimized during method development, it is important to assure that the assay is under control when the method is applied [57]. For this reason, more recent studies have required the assessment of ISR. An underlying objective is to determine whether and why the variance noted from incurred samples differs from QC samples.

Within GLP toxicology studies, a proper evaluation of ISR and accuracy needs to be performed once on each species used in toxicology studies. Consistency of the animal's genetics and environment makes it unnecessary for additional incurred sample investigations to be performed once the initial assessment has been performed. The chance of incurred sample variability is far greater in humans. Therefore, randomly

selected subjects should include a minimum of 10% of total analyzed samples. For larger studies, 5% of total sample size is allowed. Samples must be analyzed within the period of stability [12,13].

In both GLP and clinical studies, the difference between the initial and reanalyzed values should be $\leq$20% (small molecules) in at least 67% of reassayed samples. Unlike repeat analysis for anomalous results or for PK reasons, ISR is performed in singlet from selected samples about the C_{max} and in the elimination phase. It is recommended that the selected samples be representative of each study period and dose cohort. Initially, this was recommended to be referenced to the original result, assuming that it was more likely the correct value. The proper approach is to reference to the mean value:

$$\%\text{Variability} = \frac{\text{Repeat} - \text{Original}}{\text{Mean}} \times 100$$

Reference to the mean produces a consistent assessment, regardless of the order by which the two results were obtained [13]. For this reason, LIMS software, such as Watson, support either calculation.

When the criteria cannot be met, an investigation to understand the assignable cause of the irreproducibility is warranted. This will proceed from several probable causes to one assignable cause [58]. The simplest cause may be an assignable execution error, which may isolate the failure to that day's run. More complex reasons may implicate other reasons (Table 17.3). Once defined and corrected, a decision regarding reporting of data can be made. Study sample analysis cannot be accepted until the investigation is complete, documented, and appropriate follow-up actions are in place. In case of an unassignable cause, all ISR samples can be reanalyzed, along with additional samples, in two or more different batches. Results of all relevant batches should be evaluated to obtain interday ISR data. No assignable cause can require the redevelopment/validation and reanalysis of all subject samples.

TABLE 17.3 Assay Issues and Probable Causes for ISR Failures

Issue	Causes
Improper processing of study samples	Execution error (dilution, IS spiking) not seen in QC samples
Lack of rugged method	Assay not optimized and variance in study samples greater than that in QC samples
Interference from isomeric or isobaric metabolites	Metabolic specificity not fully tested
Metabolites $\rightarrow$ parent	Improper stabilization procedure
Rate of interconversion affected by the law of mass action	Interconverting forms not tested at same ratio as in study samples
Subject differences in protein binding	Extraction process not rugged
Differential recovery	Time-dependent recovery observed in study samples but not QC samples
Sample nonhomogeneity	Sampling error (improper thaw/vortex)
Endogenous interference and matrix effects	Error in defining matrix variability across populations because of insufficient lot screening

Investigation of ISR failures may implicate results, runs, studies, or assays. Understanding the cause will determine its impact and the direction taken. A unique sample may not have a reportable value. A unique population may present problems for the assay, failing study results and requiring revision/revalidation of a new assay. If the assay is implicated, it can invalidate results from other studies in which the assay was used. It is therefore critical to thoroughly test assays before use and ensure proper training of staff.

ISR results may be documented in the final bioanalytical or clinical study report and/or as an addendum to the method validation report. In selecting samples to be reanalyzed, a variety of concentrations, time points, and patient populations should be chosen, depending on the drug and its DMPK (drug metabolism and pharmacokinetic) properties. This can be a challenge when the analyst has limited information. First-in-human, proof-of-concept, special population, and bioequivalence studies are examples of studies that must include incurred sample concentration verification. Most perform ISR within all clinical studies. Study sample results obtained for establishing ISR may be used for comparison purposes and are generally not used in calculating reported sample concentrations.

To properly handle failed ISR runs, the laboratory must have an SOP on how to conduct and document investigations. The principal investigator should lead the analytical investigation and report their findings to the study director or the individual who monitors the overall conduct of the study, quality assurance, and management. Probable causes are reduced to definitive cause(s) through a series of hypothesis testing. The investigation report is issued with the study report in which the problem is described, reanalysis results are defined, and an assignable cause is identified. The identified cause serves to define preventative and corrective actions to avoid future errors.

Ruggedness testing during method development and validation can serve to reduce ISR failures. Robustness assessed by tracking assay performance across different analysts, laboratories, and studies can be invaluable in defining trends and causes for failures [59]. A critical question is to determine why there was an analytical error not uncovered by normal standard and QC acceptance criteria. Some assay-related causes of ISR failures are shown in Table 17.3. It is expected that ISR failures be fully investigated to an assignable cause and corrected, in a manner similar to assessing manufacturing failures when out-of-specification results are seen. It should be considered that imprecise assays may, on occasion, fail ISR in the same statistical manner in which one set of QC samples fail, while others pass acceptance criteria. The analyst should work to ensure that the assay can be executed more precisely before undertaking a larger ISR sampling.

17.13 MULTIPLE ANALYTE ASSAYS

It is becoming more common to expect that LC-MS assays perform multiple analyte assessments within a single run [27,48]. Cassette dosing and metabonomics are the extreme examples [60,61]. Within regulated bioanalysis, this is desired for antiviral indications where comedication drug therapy (highly active antiretroviral therapy (HAART)) is needed to overcome drug resistance. In many areas, the desire to measure parent drug and its metabolites drives multiplexed assays.

Simultaneously measuring multiple analytes increases the statistical probability of run failures. Results should not be rejected for the remaining analytes if one analyte fails. Concentrations from the first accepted run should be reported. If this analyte is repeated in simultaneous assays when analyzing for different analytes, it is not necessary to quantitate the already reported analytes. All results should be maintained and available for review.

17.13.1 Drug Metabolites

It is common to have metabolite exposures at significantly different levels than the parent drug, particularly in early clinical studies where metabolite levels are unknown. On repeat analysis of diluted ALQ parent samples, a second reportable value for the metabolite is often obtained. Source data from all acceptable runs, regardless of whether the concentrations were reported, should be retained.

MIST guidance requires one to demonstrate steady-state exposure in at least one toxicology species for all major (>10% of parent AUC) circulating human metabolites. Traditionally, this has been assessed by single dose ADME studies using radiolabeled drug. Studies such as these can be difficult and expensive to perform early in drug development. It is worth noting that the requirement to demonstrate exposure at steady state has made multiple tracer (100 nCi) dosing using AMS a viable option. Regulatory expectations are that these assessments will be completed after proof-of-concept trials but before launching larger clinical trials. This affords some time to assess steady-state exposures and nonradiolabeled approaches such as using tiered assays.

For extensively metabolized drug candidates, measuring large numbers of metabolites at steady state may be better achieved using tracer dose AMS studies. If one chooses to measure metabolites using LC-MS assays, tiered assays afford a means to make more informed decisions. In comparison to a traditional ADME study (50 μCi dose, liquid scintillation counting), multiple LC-MS assays defer synthesis of GMP (good manufacturing practice)-grade radiolabel, tissue distribution, dosimetry estimates, and the cost of a clinical ADME study. Savings by attrition of compounds that do not meet proof of concept is significant. In contrast to doing nothing until a late clinical ADME study is performed, the tiered assay approach avoids surprises and focuses resources on unanticipated, unique, or disproportionate metabolites.

Without reference standards, UV and MS responses are sometimes used as an estimate of parent drug and metabolite levels. UV response is often more consistent, particularly when its chromophore is unaffected by metabolism. MS ionization and MS/MS fragmentation have considerable variability for structurally similar (and dissimilar) metabolites. One should not assume that UV or MS response will be comparable. Assessment at high levels from *in vitro* sources can afford more security in these estimates. Relative response factors of the metabolite to parent drug are determined and used to calibrate MS response for lower level determinations. Comparisons of this type should be considered as first-tier assays.

Another approach compares response ratios across species. This assumes similar stability, recovery, matrix effects, and response proportionality in all species or minimizes them by dilution into human plasma. Interspecies differences in enzyme levels, protein binding, and phospholipid levels suggest that analyte stability, recovery, and matrix effects may vary across species. Mixing animal samples with an equal amount of human plasma compensates for these differences and has been shown to be a good procedure

by which simple response ratios afford safety margin estimates [62]. Response proportionality by testing study samples at multiple dilutions (parallelism) is also required. If studies were performed at different times and samples stored for different periods of time, long-term stability is still in question. Reanalysis of aged study samples may be used to support long-term stability. However, reference of stability to initial response rather than its nominal concentration may mask instability, as the breakdown product of one metabolite may fortify an unstable analyte. Ratio comparisons across species should be regarded as second-tier assays.

Third-tier assays generally employ calibrators that have been qualified with some limited assessment of stability. One approach previously mentioned uses metabolites from radiolabeled animal or human *in vitro* studies spiked into human plasma as standards for quantification [31]. To ensure accurate quantification, the radiolabeled profile must separate all metabolites from each other. Metabolite radioactivity is referenced to parent, which has been measured using LC-MS. Resulting curve ranges, therefore, reflect the original proportions within the spiked radiolabeled sample. Dilution in matrix affords the standard curve. The advantage of this approach is that steady-state assessments across species can be made quickly without the need to isolate or synthesize metabolites. Human metabolites that are disproportionate are readily noted. Those metabolites that are insignificant or for which there is adequate safety coverage may need little further investigation. It should be noted that similar approaches are possible using nonradiometric detection, provided response is specific only to metabolites and proportional to their concentration. Unfortunately, inductively coupled plasma-mass spectrometry (ICP-MS), chemical reaction interface for mass spectrometry (CRIMS) and N or S-chemiluminescence detection suffer from either high background in biological matrices or an inability to detect low background elements such as fluorine [63–65]. This approach, however, requires that the metabolite is observed in animals or in human *in vitro* sources.

A second, more time-consuming approach involves isolation of individual metabolites before NMR identification and quantification [29,30]. This approach is closer to the traditional method of isolating metabolites, synthesizing them, and assessing their purity. The difference is that the metabolite isolated from biological sources need not achieve a level of purity expected by synthesis. Only microgram levels are needed, and although the isolated metabolite should be the major component, it need not be high in purity. As NMR response is independent of structure, accurate quantification is achieved by direct measurement of a metabolite resonance using a parent peak for calibration. In contrast, synthetic reference standards are generally qualified by difference, subtracting impurity identified in their preparation.

Since cross-species comparisons reference human-to-animal exposure, the exact assessment of purity is not critical. That said, the direct determination is likely more accurate when the compound cannot be either synthesized or isolated in high purity. Limited supplies will make it difficult to characterize impurities, so a direct measurement will be more accurate. Likewise, the inclusion of a reference standard provides much greater accuracy than just measuring response. Furthermore, the direct measurement approach using NMR should provide both the confidence needed for structural integrity and purity assessment for its use in a regulatory assay.

In combination with strategies for biologically sourcing metabolites and qualifying them without a certificate of analysis, tiered assays can provide a reasonable estimate of steady-state exposures. Samples should be stored at -70°C and analyzed as soon as

received. Whether this is achieved using single or multiple assays is dependent on the extent and diversity or metabolism, required sensitivity and dynamic ranges, and risks one is willing to take by trying to achieve too much within a single assay. Analysts and development teams should consider whether a metabolite is worthy of GLP bioanalysis at its present stage of drug development or whether a tiered approach would be better.

Human metabolites for which there is abundant safety coverage are of less concern and support a tiered determination. The MIST guidance should not replace structure–reactivity assessments. Metabolites not expected to be active or reactive, such as sulfate or ether glucuronides, pose less concern [66]. Alternatively, low levels of highly reactive metabolites may not be seen in plasma, as they do not escape the liver in significant abundance. Accurate measurement of a trapped metabolite or both precursor and product metabolites in plasma may be required. Each metabolite should be considered on a case-as-case basis. Having an accurate assessment is critical when metabolism-mediated toxicity is observed [67].

Once the human metabolism is known, a solid decision can be made regarding what assay(s) would be effective in supporting clinical drug development. To allow proper assessment of PK–PD, parent and significant human metabolites should be measured to regulatory standards. It may be worthwhile to include abundant metabolites into this assay, particularly if they are active, potentially toxic (structural alerts), or implicated in drug clearance. Decisions regarding assays for toxicology studies in animals include safety concerns in humans.

17.13.2 Drug Interaction Studies

DDI is a major reason for toxicity, particularly when new drugs are used in larger, more diverse populations. New chemical entities often clear the body as metabolites or through transporters, both of which are subject to inhibition or induction by other drugs or their metabolites. In addition, the presence of genetic polymorphism and disease as well as age- or gender-related differences in enzyme or transporter function can serve to reduce the normal safety margin. The previous FDA Guidance on DDI studies has focused on *in vitro–in vivo* correlations, providing guidance on what studies are needed in clinical drug development [68,69]. More recent publications have addressed both CYP-mediated and transporter-mediated drug interactions [70,71].

When desired, it is possible to perform a more general screening of new drug candidates using a cocktail of marketed drugs that serve as cytochrome P450 probe substrates. As many as 10 analytes (parent and metabolite for 5 probe substrates) are measured in this coadministration study. Common substrates in DDI cocktails include warfarin (2C9), dextromethorphan (2D6), omeprazole (2C19), midazalom (3A4), and caffeine (1A2, NAT-2, XO). Other combinations have used debrisoquin (2D6), (*S*)-mephenytoin (2C19), chlorzoxazone (2E1), caffeine (1A2), and dapsone (3A4). Flurbiprofen has been added to cover 2C9. Other cocktails include mixtures of preferred probe substrates that assess clinical relevance to the drug under development [72,73]. This approach could find more utility, provided a common set of probe substrates and metabolites were used. Mixing of probe substrates and metabolites requires more extensive assay development and validation for each DDI study.

Measuring a metabolite that is specific to one biotransformation pathway provides a more powerful assessment of a drug interaction than just measuring a change in parent drug exposure. Unfortunately, the interaction of large and small molecules is

less predictable than small-molecule interactions. Small-molecule interactions mostly occur as a result of liver enzyme or transporter inhibition and are well predicted from *in vitro* studies. Large–small molecule interaction can be target-related and noted from enhanced or reduced pharmacological effect [74]. Regardless, most drug development programs now include an assessment of the effects of biologics on the systemic levels of small-molecule drugs.

17.14 ASSAY TRANSFERS AND CHANGES

There can be multiple opportunities to transfer assays. Each of these affords a chance to improve an assay. Assay transfer within a laboratory or across laboratories involves an assessment of its equivalence following the change. Depending on the nature of the change, this may involve either a partial or full validation. Transfers do not require cross-validation or conformance testing. However, testing at least QC samples from the other laboratories may expose a bias. Inclusion of incurred samples may expose issues not seen using just QC samples. Laboratories need to demonstrate ISR many times within clinical development, so requiring a similar test at assay transfer would be a wise investment. Cross-validation using incurred samples is only required when two different assays are used within the same study. However, at New Drug Application (NDA) filing, an assessment of exposures across populations is made, so assays are expected to give equivalent results in all studies.

At the discovery-to-GLP transfer, one should take every opportunity to learn from discovery studies. MS optimization should test ionization, instrument tuning, sensitivity, and assay range. Often, a generic extraction and fast gradient separation suitable for discovery needs revision to meet regulatory expectations. Automated screening of LLE (liquid–liquid extraction) or SPE procedures can test diverse conditions for high recovery and minimum matrix effects [75]. Mobile phase and LC column screening should test gradient or isocratic conditions using various mobile phases and columns. Attention to the separation of any interference, carryover, peak shape, and MS sensitivity should be made. MS sensitivity is affected by the separation for several reasons. ESI is concentration dependent, and mass flux to the source is impacted by peak shape and resolution. Efficient drying of ESI droplets is aided by reducing surface tension, so better response is achieved when using a higher percentage of organic in the mobile phase. Having the right pH or ionizing agent (e.g., ammonium) is also critical to ionization. For absorptive compounds, through-column recovery may be low. While calibrators may compensate for reduced LC recovery, the net effect is poorer sensitivity. It is also common to substitute a better, more abundant analog or preferably stable label IS at this phase of method development.

At the IND toxicology-to-clinical phase, assay sensitivity and a wide dynamic range are critical as one prepares for SAD (single ascending dose) and MAD studies in man. Achieving a high sensitivity assay may require larger volumes and the use of a more selective extraction or exploring alternative ionization or derivatization approaches. The addition of active metabolites to better assess PK–PD is a common change. At this time, it is unwise to include too many metabolites until samples from the top dose of the MAD study are profiled. Profiling and tiered assays can provide estimates as to whether these metabolites are unique or disproportionate and therefore require

long-term monitoring using a regulated assay. This approach will minimize changes to the clinical assay throughout drug development. As discussed previously, a human urine assay is generally needed.

When moving further to proof of concept, assays in other fluids or tissues may be needed to demonstrate penetration to the site of action or toxicity. Site-of-action questions will generally require analysis only in human and the pharmacology species. Toxicity questions will include human and the most sensitive safety species. In many cases, a representative human sample is unavailable. For human lung, bronchoalveolar lavage and macrophage exposure can serve as its surrogate. Some virology indications are supported by showing penetration into PBMCs. This is particularly important when drug is known to be actively transported or limited by intracellular metabolism. There is limited guidance on the extent of validation required for a penetration study. If limited to a single matrix and critical to explaining efficacy or toxicity, a full validation including stability assessments should be considered. For tissues, stability assessments may be limited to homogenates. This is generally sufficient for toxicology studies, as drug safety sites often can perform homogenization before freezing tissue samples.

Whenever assays are transferred across different companies, such as when a pharmaceutical company transfers an assay to a contract research organization, there is risk that any change could impact assay performance. The dilemma is whether one should allow changes to better optimize an assay or stay with the present assay. The transfer of knowledge and good communication are critical. Assessments of accuracy, precision, and failure rates provide metrics on its success. To confirm a lack of bias, the exchange and analysis of both QC and incurred samples can demonstrate equivalence before starting a new study. If additional tests are required beyond what is normally covered by the SOP, preparing and executing a validation plan is essential.

17.15 DOCUMENTATION

When preparing analytical and validation reports, a summary table of all analytical runs should list runs with run identification, dates of analysis, whether runs passed or failed, reasons for any failures, and, for analytical runs, any deviations from the validated method. QC data from validation runs that only failed to meet QC acceptance criteria with no assignable cause for failure should be included in the precision and accuracy estimation. Deviations from SOPs and assay procedures and significant unexpected events should be identified and their impact assessed. This includes any defined outliers [12]. For those performing large numbers of studies, having a LIMS is a critical component to ensure consistency in documentation and to manage the time associated with these tasks.

Source data documentation should include the conditions of use. For instance, stability determinations during method validation should record experimental conditions such as storage temperature and duration. Materials should be traceable to source and date of collection. Modification of calibration response and QC levels should be documented with sufficient detail to demonstrate that the changes were justified or followed established procedures. For chromatographic methods, source documentation should include original and reintegrated chromatograms for accepted runs, along with the reason for changing integration parameters across a run or for individual samples within a run. Electronic audit trails that record changes to integration parameters must be enabled [12].

A total of 5% of all chromatograms from randomly selected subjects, including QC samples and standards, must be submitted with an NDA and/or ANDA (Abbreviated New Drug Application). For bioequivalence studies, 20% of chromatograms from serially selected subjects are submitted.

Final reports for either validation or study sample analysis must include a complete account of the performance of the method. This includes a tabular listing of the actual QC results from all runs during method validation and accepted runs during study sample analysis. A table listing reassayed samples, reason for reanalysis, and values for the original, reassay, and final reportable result should be included in the final report.

Drug concentration data from the rejected runs need not be included in the final report; however, a brief description of the reasons and a tabular listing of rejected runs should be provided. Compliance has the added value that thorough documentation at the end of each study makes work easy when the common technical document (CTD) is assembled.

17.16 INSTRUMENT QUALIFICATION

Qualification of instrumentation and computer validation are beyond the scope of this chapter. Suffice it to say that the proper installation and operational and performance qualification of all instrumentation are required before performing regulated bioanalysis [76–79]. This responsibility is shared among the vendor, analysts, and informatics organizations. Testing and validation of functionality unique to your laboratory requirements will not have been performed by the vendor and therefore must be done. Design qualifications tend to be the domain of the vendor, although this aspect can be assessed during the vendor audit and at the time of instrument selection. User specifications and a validation plan should be developed. From vendor assessments, system functionality can be verified and any gaps identified. Consider customization when configuring and testing options. UAT should include a traceability matrix to regulations. Failures in executing UAT scripts are categorized as to their severity and decisions made for any procedural or software changes. Results and all reports must be thoroughly reviewed before the production release memo is issued. SOPs, postproduction, validation review, change management, disaster recovery, deviation, and exception documentation are required.

17.17 EXPLORATORY CLINICAL STUDIES

Many organizations have sufficient discovery or preclinical resources and confidence in animal studies that would predict success in the clinic. Demonstrating proper exposure at a therapeutic dose or target engagement in the clinic is expected for IND candidates. When unsure of success, eIND studies allow the assessment of new chemical entities that have undergone limited safety testing [80]. Subtherapeutic doses or microdoses afford the opportunity to determine ADME properties using AMS or target engagement using a positron emission tomography (PET) ligand [81]. A microdose is defined as $<$1/100th of the dose of a test substance calculated (based on animal data) to yield a pharmacologic effect of the test substance with a maximum dose of $\leq 100\,\mu g$.

For some, the eIND presents a fast-to-fail decision. There is always the concern that DMPK properties at very low doses are not reflective of therapeutic doses because of dissolution rate-limited absorption or saturation of clearance [82]. Having clarity on the biopharmaceutics is therefore fundamental to understanding whether one or several new chemical entities are good candidates for microdosing. Not having a good *in vitro–in vivo* correlation and poor allometric scaling of the PK in animal species could be justification.

If microdose PK is the only desired endpoint, cold dosing of microgram quantities in humans and LC-MS measurements of picogram per milliliter concentrations in plasma are alternative means to estimate human exposures [83]. In a preclinical test, proportionality for several drugs in rats was considered as validation of a general methodology to support microdosing in humans [84]. Sprague-Dawley rats were given oral doses of 0.167, 1.67, 16.7, 167, or 1670 mg/kg. LC-MS provided sufficient sensitivity to study the PK of antipyrine, carbamazepine, atenolol, and digoxin, but not metoprolol. Dose proportionality was achieved between a microdose and up to 1000-fold higher doses for antipyrine, carbamazepine, and digoxin. State-of-the-art MS (API 5000) instrumentation was required. There was limited sample cleanup and no derivatization. The LLOQs ranged from 5 to 20 pg/mL. This illustrates the potential to quickly adapt existing LC-MS assays to an eIND study. Others have followed with examples of the potential of microdosing to obtain an early assessment of ADME properties [85].

High sample enrichment and derivatization may be needed to achieve microdosing objectives. For some small molecules and most peptides, immunoprecipitation can be used to enrich a larger volume of sample before analysis [86]. If the capture antibody from an enzyme-linked immunosorbent assay (ELISA) is used, the resulting LC-MS assay will share its epitope specificity. More details are presented later in the analysis of larger molecules. Molecularly imprinted polymers, sometimes referred to as *plastic antibodies*, are an alternate means to enrich samples [87]. They are more readily produced than antibodies but have considerable limitations. Derivatization generally employs means to make molecules more thermally stable or readily ionized in APCI or by attaching ionized surfactants in ESI [88,89]. In either approach, one should test derivatization conditions to ensure that they are robust and carried to completion. Limiting reagents or conditions can generate differences across samples.

17.18 LARGER MOLECULES

Peptides, small proteins, and oligonucleotides are representative of mid-size molecules (1–10 kDa) that can be directly measured using LC-MS to low nanomolar or subnanomolar levels. Van den Broek [90] has reviewed the use of LC-MS methodology to perform quantitative bioanalysis of peptides in biological fluids. Noted are the common requirements to add displacement reagents (similar peptides) or other proteins (serum albumin) to compete with binding sites and to add acid, solvents, or surfactants to overcome nonspecific binding. Immunodepletion of most abundant plasma proteins is also mentioned as a means to enrich samples. If time allows, developing an antibody to immunocapture the desired peptide would be preferred. In an example of derivatization, an octyl quaternary ammonium group was added to enhance ESI properties [89]. The addition of both a charged and hydrophobic group by derivatization aids

by enhancing its surface properties and ion yield. Reporting of both extraction and recovery through immunoprecipitation and derivatization from the low and high QC samples in the validation would be expected. Yields should be sufficiently high in both processes so as to avoid assay variability.

The analysis of insulin using LC-MS illustrates its value both as a biomarker and, when dosed, for its PK [91]. Radioimmunoassays (RIAs) and ELISA of insulin often are nonspecific across higher species, making it difficult to separate endogenous animal insulin from dosed human insulin in toxicology studies. Dosing of human insulin reduces endogenous animal insulin, so measuring a composite profile with a nonspecific RIA is problematic. Predose and insulin precursor C-peptide levels can be determined, but a more direct measurement with a specific assay affords one the chance to measure both drug levels and the effect.

While peptides may have picomolar potency and may require high sensitivity assays in clinical studies, toxicology studies may be supported using far simpler LC-MS assays. Hu and Kamberi [92] reported the analysis of apolipoprotein A-1 mimetic peptide D-4F in rabbit plasma from 0.02 to 40 μg/mL. When using a mixture of LC-MS and immunoassays in toxicology and clinical studies, one should have a clear understanding of assay differences to assess safety margins. The same can be said of discovery-to-preclinical bioanalysis when assessing the minimum effective dose. Assay bias should be assessed to ensure adequate safety margins and human dose selection.

Cross-validation of ligand-binding and LC-MS assays using both QC and study samples can help to illustrate bias. Although this is not required within the guidance, it should be considered when making assessments in large molecules. Larger molecules have tertiary and quaternary structures that may be critical to ligand-binding assays but unrecognized by LC-MS. If required for potency, immunoassay could be the better choice. Likewise, differences in specificity may be exposed when comparing assays. LC-MS requires extraction, so total drug is measured, whereas some immunoassays measure only free drug levels. Differences can be considerable.

Plotting study sample results for an immunoassay on one axis and LC-MS results on another should yield a correlation. There are no *a priori* acceptance criteria. The ideal correlation is a 45° line (slope = 1) that would indicate perfect agreement between the assays. The best fitting line will display the actual relationship. A comparison of precision between the assays can be done by computing the standard deviation of replicate results for each sample. Comparison of individual and grand means can be done using analysis of variance.

Therapeutic oligonucleotides have been under development for two decades [93]. Oligonucleotides are a special challenge and present their own stability and sensitivity problems. Hybridization-based ELISA methods generally afford the required speed and sensitivity for these potent drugs. Identification of individual metabolites, including those generated from 3′- or 5′-exo/endonucleases, is generally not possible using ELISA. While LC-MS may afford a more direct means to determine the metabolism of oligonucleotides, ionization efficiency in negative ESI is often limited and variable [94]. Matching chromatographic separations to MS detection is also a challenge. Reducing and controlling the formation of salt adducts through addition of bases such as triethylamine are critical to successful analysis. At best, low nanomolar concentrations have been measured [95]. Besides the parent drug, the 3′N-1, 5′N-1, 5′N-2, and 5′N-3 metabolites were also capable of being measured in rat plasma.

These types of separations require system suitability samples to be analyzed both before and after runs to ensure that separation of any critical pairs and MS sensitivity have been maintained. It is also important to understand the activity of any metabolite. Any multiplexed assay for oligonucleotides should be based on a strategy that is similar to small-molecule drug metabolites. Significant circulating active metabolites in humans should be included with the parent assay to assess PK–PD. Hopefully, this can be achieved within a single assay. It may be impossible to source others, so a tiered strategy may be needed. Particular attention to off-target toxicity is needed, as the similarity of oligonucleotide metabolites to parent drug may force one to consider regulatory assays for all metabolites.

Differences between present acceptance rules (15/20% for small vs 20/25% for large) have resulted in some confusion when mid-size molecules are measured. For instance, when oligonucleotides or peptides are measured using LC-MS, should acceptance criteria be tighter than when measured using immunoassay? As peptides and oligonucleotides are typically manufactured without heterogeneity, tighter standards may be applied. However, both an ELISA and an immunocapture LC-MS assay may share a reagent expressed with high variability (e.g., polyclonal antibody). A reasonable approach might be to target ±15% but to allow the acceptance criteria to be adjusted up to ±20%, contingent on the accuracy and precision noted within the validation.

New approaches to deliver drugs, such as ADCs, are good examples of where the released peptide drug measured using LC-MS may be measured to the same rules as its parent biologic. In many cases, the parent ADC is measured using ELISA and results would be accepted with an accuracy and precision of 20% (25% at LLOQ). However, one could equally make the case that a small peptide measured by LC-MS as a protein surrogate should be determined to small-molecule rules (15% accuracy and precision, except 20% at LLOQ).

Moving to larger molecules (biologics) requires careful consideration of whether a highly specific LC-MS assay is useful. If ELISA measures all active components, a functional assay should be the primary measurement. What role does a peptide-as-protein surrogate assay then play in drug development? When high affinity antibodies are difficult to obtain, there may be good reason for LC-MS detection. Its higher specificity can compensate for a less selective capture reagent.

Proteins are not generally directly amenable to LC-MS bioanalysis. This is particularly true for proteins that have undergone extensive posttranslational glycosylation or to which polyethylene glycol (PEG) has been added. Proteins, following their isolation can be reduced in size using proteolysis and measured by LC-MS. Tenecteplase (MW 58,777 Da) has been determined in rat plasma in the range of 5–125 μg/mL [96]. Plasma samples were reduced using dithiothreitol and reacted with iodoacteamide before overnight trypsin digestion. LC-MS analysis of three tryptic peptides served as surrogate for the protein drug. This relatively simple procedure afforded sufficient sensitivity to determine the PK in rat. While the LLOQ of the ELISA was much lower, this approach will be of increasing use to speed early drug development.

The LC-MS approach is time consuming but can provide detailed information of amino acid changes that are unrecognized by an immunoassay. For instance, one can interrogate whether a specific Asn residue has been hydrolyzed to Asp and measure its amount in plasma. This is a common degradation of peptide drugs, including insulin. Rearrangement can also occur during this process, which would require chromatographic separation. The LC-MS assay could provide complementary information on

impurities, degradants, and metabolites that cannot be determined using an immunoassay.

Enfuvirtide (MW 4492 Da) and its deamidated metabolite M-20 were determined in human plasma [97]. The assay range for parent drug was 10–2000 ng/mL, while that for M-20 was 10–500 ng/mL. A deuterated stable label was added as IS before protein precipitation. The resulting supernatant was directly analyzed using LC-MS. The same drug and its M-20 metabolite were measured in a similar assay [98]. Deuterated stable label IS was added before chymotrypsin digestion. LC-MS analysis of four smaller peptides was shown over the 100–10,000 ng/mL range. Since the peptide drug was of sufficiently low molecular weight, either a direct or indirect analysis under low resolution MS/MS conditions afforded an accurate measurement of the separated drug and M-20 metabolite. The easier method would avoid proteolysis; however, this assay was an order of magnitude less sensitive.

Blending of exposure data across studies and species from multiple LC-MS and ligand-binding assays will be of increasing importance in the drug development of peptides, oligonucleotides, and protein therapeutics. It will be critical to demonstrate adequate analytical figures of merit for stability, recovery, matrix effects, and specificity. A fundamental understanding of the differences in these assays through appropriate cross-validation must also be obtained.

17.19 DRIED BLOOD SPOTS

DBSs have been infrequently used for many years in the field of bioanalysis. Over the past four years, there has been considerable interest, largely driven by Glaxo Smith Kline, in the use of DBSs instead of plasma [99,100]. There are numerous advantages, particularly from the *in-life* portion of the study. For rodent studies, less blood (15 μL) is required, which reduces the numbers of animals. Toxicology of satellite animal groups can be eliminated. Repetitive sampling from the same mouse also reduces interanimal variability in PK studies. In the clinic, labor-intensive plasma harvesting is eliminated. Needlestick bleeds replace indwelling catheters. Likewise, the need to store and ship samples at reduced temperatures is eliminated. The overall effect is to reduce errors at remote clinics, allowing higher quality from global studies.

The challenge becomes one for the bioanalyst. The analyte must be well recovered from filter paper without the introduction of matrix effects. As only a fraction of the total blood spot is used, the punch sampling must be sufficiently large to ensure the required LLOQ. Reemergence of DBSs has been aided by the improved performance of LC-MS systems. Owing to their sensitivity requirements, highly potent drugs may not be amenable to DBSs. However, there are a considerable number of drugs that require no greater than low nanogram per milliliter detection sensitivity, a target achievable for many DBS applications. High affinity enrichment and smaller-volume analysis including the use of capillary LC-MS may assist in obtaining additional detection sensitivity. Sampling across the blood spot must be consistent to ensure a reproducible sample since aliquoting a precisely measured volume at collection is not done. At present, automation is limited. There is considerable effort by automation companies to provide more innovative, off-the-shelf solutions in the near future.

As more investigators explore the potential of DBSs in clinical studies, there will undoubtedly be regulatory bioanalytical guidance that comes forth in the general area

of microsampling. Many are performing qualifying studies in which blood, plasma, and DBSs together are sampled and analyzed. When corrected for hematocrit and blood–plasma partitioning, the PK would be expected to be equivalent. This type of test is likely needed to bridge data for a new chemical entity that is moving through drug development.

17.20 CONCLUSIONS

The nature of regulated bioanalysis is ever changing. New methodology, technology, and novel drug candidates challenge past practices. Best policies from GMP and GCP (good clinical practice) are being incorporated into GLP. Most notable has been the use of ISR to confirm or refute original results and the use of out-of-specification procedures to investigate the causes of failed runs [101]. Although this has placed additional costs on the industry, it has greatly improved the quality of bioanalytical data. No longer does the first reportable result go unchallenged.

As a result of the Third Crystal City Bioanalytical Workshop, common procedures were adopted for validation and sample analysis of both large and small molecules, regardless of the analytical technology. These changes in regulated bioanalysis were greatly needed to assist in measuring more complex drug candidates. Products made by synthetic routes are integrated with those from biological sources, including modified peptides, proteins, and antibodies. LC-MS and ligand-binding assays share drug development roles, defining PK–PD and assessing drug safety. As such, it is likely that future guidance may also better integrate biomarker assays into regulated bioanalysis.

Global drug development has brought forward new cultures and perspectives into our science. Adaptive clinical trial design put more emphasis on bioanalysis to escalate doses and biomarker assays to define proof of concept. The next generation of regulated bioanalysis will see new challenges from evolving science and technology that will further shape work practices and standards.

REFERENCES

1. Karnes HT, Shiu G, Shah VP. Validation of bioanalytical methods. Pharm Res 1991; 8:421–416.
2. Shah VP, Midha KK, Dighe S, *et al*. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Conference report. Pharm Res 1992;9:588–592.
3. Hartmann C, Smeyers-Verbeke J, Massart DL, *et al*. Validation of bioanalytical chromatographic methods. J Pharm Biomed Anal 1998;17:193–218.
4. Causon R. Validation of chromatographic methods in biomedical analysis. Viewpoint and discussion. J Chromatogr B 1997;689:175–180.
5. Smith WC, Sittampalam GS. Conceptual and statistical issues in the validation of analytic dilution assays for pharmaceutical applications. J Biopharm Stat 1998;8:509–532.
6. International Conference on Harmonization (ICH). Validation of analytical methods: methodology, ICH Q2 A. 1994.
7. International Conference on Harmonization (ICH). Validation of analytical methods: methodology, ICH Q2 B. 1994.
8. Shah VP, Midha KK, Findlay JW, *et al*. Bioanalytical method validation—a revisit with a decade of progress. Pharm Res 2000;17:1551–1557.
9. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry, bioanalytical method validation. 2001.

10. Findlay JW, Smith WC, Lee JW, *et al*. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. J Pharm Biomed Anal 2000;21:1249–1273.
11. DeSilva B, Smith W, Weiner R, *et al*. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. Pharm Res 2003;20:1885–1900.
12. Viswanathan CT, Bansal S, Booth B, *et al*. Workshop/Conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. AAPS J 2007;9:E30–E42.
13. Fast DM, Kelly M, Viswanathan CT, *et al*. 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;11:238–241.
14. Van Eeckhaut A, Lanckmans K, Sarre S, *et al*. Validation of bioanalytical LC–MS/MS assays: evaluation of matrix effects. J Chromatogr B 2009;877:2198–2207.
15. European Medicines Agency. Concept paper/recommendations on the need for a (CHMP) guideline on the validation of bioanalytical methods. 2008.
16. Agência Nacional de Vigilância Sanitária (ANVISA). Guide for validation of analytical and bioanalytical methods. 2003.
17. Savoie N, Garofolo F, van Amsterdam P, *et al*. 2009 White paper on recent issues in regulated bioanalysis from the third calibration and validation group workshop. Bioanalysis 2010;2:53–68.
18. Lausecker B, van Amsterdam P, Brudny-Kloppel M, *et al*. European bioanalysis forum and the way forward toward harmonized regulations. Bioanalysis 2009;1:873–875.
19. Bolton S. and Bon C. Linear regression and correlation. In: Swarbrick J editor. Pharmaceutical statistics: practical and clinical application. 5th ed. New York (NY): Informa Healthcare; 2010. pp. 163–166.
20. Heller D. Ruggedness testing of quantitative atmospheric pressure ionization mass spectrometry methods: the effect of co-injected matrix on matrix effects. Rapid Commun Mass Spectrom 2007;21:644–652.
21. European Medicines Agency. Guidelines on the evaluation of control samples in nonclinical safety studies: checking for contamination with the test substance. 2006.
22. Zeng H, Wu JT, Unger SE. The investigation and the use of high flow column-switching LC/MS/MS as a high-throughput approach for direct plasma sample analysis of single and multiple components in pharmacokinetic studies. J Pharm Biomed Anal 2002;27:967–982.
23. Ji A, Saunders JP, Wadgaonkar ND, *et al*. A novel antibiotic bone assay by liquid chromatography/tandem mass spectrometry for quantitation of tigecycline in rat bone. J Pharm Biomed Anal 2007;44:970–979.
24. Wu D, Paul DJ, Zhao X, *et al*. A sensitive and rapid liquid chromatography-tandem mass spectrometry method for the quantification of the novel neurokinin-1 receptor antagonist aprepitant in rhesus macaque plasma, and cerebral spinal fluid, and human plasma with application in translational NeuroAIDs research. J Pharm Biomed Anal 2009;49:739–745.
25. Wu Y, Wen J, Fan G, *et al*. Development and validation of a selective and robust LC-MS/MS method, employing liquid-liquid extraction based on 96-well format plates, for high-throughput quantification of docetaxel in small plasma samples. Instrum Sci Technol 2009;37:1–22.
26. Shi G, Wu JT, Li Y, *et al*. Novel direct detection method for quantitative determination of intracellular nucleoside triphosphates using weak anion exchange liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom 2002;16:1092–1099.
27. Elens L, Veriter S, Yombi JC, *et al*. Validation and clinical application of a high performance liquid chromatography tandem mass spectrometry (LC-MS/MS) method for the quantitative determination of 10 anti-retrovirals in human peripheral blood mononuclear cells. J Chromatogr B 2009;877:1805–1814.

28. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry, safety testing of drug metabolites. 2008.
29. Espina R, Yu L, Wang J, *et al*. Nuclear magnetic resonance spectroscopy as a quantitative tool to determine the concentrations of biologically produced metabolites: implications in metabolites in safety testing. Chem Res Toxicol 2009;22:299–310.
30. Dear GJ, Roberts AD, Beaumont C, *et al*. Evaluation of preparative high performance liquid chromatography and cryoprobe-nuclear magnetic resonance spectroscopy for the early quantitative estimation of drug metabolites in human plasma. J Chromatogr B 2008;876:182–190.
31. Zhang D, Raghavan N, Chando T, *et al*. LC-MS/MS-based approach to obtaining exposure estimates of metabolites in early clinical trials using radioactive metabolites as reference standards. Drug Metab Lett 2007;1:293–298.
32. Petucci C, Lloyd TL, Harris HA, *et al*. Trace LC/MS/MS quantitation of 17β -estradiol as a biomarker for selective estrogen receptor modulator activity in the rat brain. J Mass Spectrom 2009;44:65–71.
33. Ocana MF, Neubert H. An immunoaffinity liquid chromatography-tandem mass spectrometry assay for the quantification of matrix metalloproteinase-9 in mouse serum. Anal Biochem 2010;399:202–210.
34. Rohrer R. Antibody drug conjugates: potent weapons for the oncology arsenal. Chem Today 2009;27:56–60.
35. Bolton S. and Bon C. Transformations and outliers. In: Swarbrick J editor. Pharmaceutical statistics: practical and clinical application. 5th ed. New York (NY): Informa Healthcare; 2010. pp. 250–252, 477.
36. King R, Bonfiglio R, Fernandez-Metzler C, *et al*. Mechanistic investigation of ionization suppression in electrospray ionization. J Am Soc Mass Spectrom 2000;11:942–950.
37. Matuszewski B, Constanzer M, Chavez-Eng C. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem 2003;75:3019–3030.
38. Bennett P, Van-Horne KC. Identification of the major endogenous and persistent compounds in plasma, serum, and tissue that cause matrix effects with electrospray LC/MS techniques. Proceedings of the 2003 AAPS annual meeting and exposition. Salt Lake City (UT): 2003 Oct.
39. Little J, Wempe M, Buchanan C. Liquid chromatography-mass spectrometry/mass spectrometry method development for drug metabolism studies: examining lipid matrix ionization effects in plasma. J Chromatogr B 2006;833:219–230.
40. Kebarle P, Teng L. From ions in solution to ions in the gas phase. Anal Chem 1993; 65:972A–986A.
41. Covey T, Lee E, Henion J. High-speed liquid chromatography/tandem mass spectrometry for the determination of drugs in biological samples. Anal Chem 1986;58:2453–2460.
42. Robb D, Covey T, Bruins A. Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. Anal Chem 2000;72:3653–3659.
43. Xia YQ, Jemal M. Phospholipids in liquid chromatography/mass spectrometry bioanalysis: comparison of three tandem mass spectrometric techniques for monitoring plasma phospholipids, the effect of mobile phase composition on phospholipids elution and the association of phospholipids with matrix effects. Rapid Commun Mass Spectrom 2009;23:2125–2138.
44. Shi G. Application of co-eluting structural analog internal standards for expanded linear dynamic range in liquid chromatography/electrospray mass spectrometry. Rapid Commun Mass Spectrom 2003;17:202–206.
45. Plebani M, Lippi G. Hemolysis index: quality indicator or criterion for sample rejection. Clin Chem Lab Medicine 2009;47:899–902.
46. Hughes NC, Bajaj N, Wong J, *et al*. Assessing the matrix effects of hemolyzed samples in bioanalysis. Bioanalysis 2009;1:1057–1066.

47. Wu JT, Zeng H, Qian M, *et al*. Direct plasma sample injection in multiple-component LC-MS-MS assays for high-throughput pharmacokinetic screening. Anal Chem 2000;72:61–67.
48. Shi G, Lloyd TL, Sy SKB, *et al*. Simultaneous quantification of seven active metabolites of roxifiban in human plasma by LC/MS/MS in the presence of an interfering displacer at millimolar concentrations. J Pharm Biomed Appl 2003;31:937–951.
49. Brockman AH, Hatsis P, Paton M, *et al*. Impact of differential recovery in bioanalysis: the example of bortezomib in whole blood. Anal Chem 2007;79:1599–1603.
50. Zeng W, Musson DG, Fisher AL, *et al*. A new approach for evaluating carryover and its influence on quantitation in high-performance liquid chromatography and tandem mass spectrometry assay. Rapid Commun Mass Spectrom 2006;20:635–640.
51. Nowatzke W, Woolf E. Best practices during bioanalytical method validation for the characterization of assay reagents and the evaluation of analyte stability in assay standards, quality controls, and study samples. AAPS J 2007;9:E117–E122.
52. Ji A, Jiang Z, Livson Y, *et al*. Challenges in urine bioanalytical assays: overcoming non-specific binding. Bioanalysis 2010;2:1573–1585.
53. Andrew V, Stachulski AV, Harding JR, *et al*. Acyl glucuronides: biological activity, chemical reactivity, and chemical synthesis. J Med Chem 2006;49:6931–6945.
54. Xue YJ, Liu J, Simmons NJ, *et al*. Separation of a BMS drug candidate and acyl glucuronide from seven glucuronide positional isomers in rat plasma via HPLC with MS/MS detection. Rapid Commun Mass Spectrom 2006;20:1776–1786.
55. Hayes R. Leveraging advances in HPLC to maximize DMPK data quality. In: Muddiman D, Cohen L, editors. Volume 5, Bioanalytical approaches: LC-MS, NMR, isotope methods, spectroscopy, ligand-binding and other analytical chemistry methods. New York: John Wiley and Sons; 2011.
56. Chang MS, Kim EJ, El-Shourbagy TA. A novel approach for in-process monitoring and managing cross-contamination in a high-throughput high-performance liquid chromatography assay with tandem mass spectrometric detection. Rapid Commun Mass Spectrom 2006;20:2190–2200.
57. Jemal M, Ouyang Z, Powell ML. A strategy for a post-method validation use of incurred biological samples for establishing the acceptability of a liquid chromatography/tandem mass-spectrometric method for quantitation of drugs in biological samples. Rapid Commun Mass Spectrom 2002;16:1538–1547.
58. Bakhtiar R, Majumdar TK. Tracking problems and possible solutions in the quantitative determination of small molecule drugs and metabolites in biological fluids using liquid chromatography–mass spectrometry. J Pharmacol Toxicol Methods 2007;55:227–243.
59. Merves ML, Goldberger BA. Quality assurance, quality control and method validation in chromatographic applications. In: Bertholf RL, Winecker RE, editors. Chromatographic methods in clinical chemistry and toxicology. New York: John Wiley & Sons; 2007. pp. 1–14.
60. Berman J, Halm K, Adkison K, *et al*. Simultaneous pharmacokinetic screening of a mixture of compounds in the dog using API LC/MS/MS analysis for increased throughput. J Med Chem 1997;40:827–829.
61. Wilson ID, Plumb R, Granger J, *et al*. HPLC-MS-based methods for the study of metabonomics. J Chromatogr B 2005;817:67–76.
62. Gao H, Deng S, Obach RS. A simple liquid chromatography-tandem mass spectrometry method to determine relative plasma exposures of drug metabolites across species for metabolite safety assessments. Drug Metab Dispos 2010;38:2147–2156.
63. Smith CJ, Shillingford S, Edge AM, *et al*. Quantification of the *in vitro* and *in vivo* metabolic fates of 2-, 3- and 4-bromobenzoic acids using high temperature LC coupled to ICP-MS and linear ion trap MS. Chromatographia 2008;67:673–678.

64. Goldthwaite CA, Hsieh FY, Womble SW, *et al*. Liquid chromatography/chemical reaction interface mass spectrometry as an alternative to radioisotopes for quantitative drug metabolism studies. Anal Chem 1996;68:2996–3001.
65. Taylor EW, Jia W, Bush M, *et al*. Accelerating the drug optimization process: identification, structural elucidation and quantification of *in vivo* metabolites using stable isotopes and LC/MS^n and the chemiluminescent nitrogen detector. Anal Chem 2002;74:3232–3238.
66. Humphreys WG, Unger SE. Safety assessment of drug metabolites: characterization of chemically stable metabolites. Chem Res Toxicol 2006;19:1564–1569.
67. Mutlib AE, Gerson RJ, Meunier PC, *et al*. The species-dependent metabolism of efavirenz produces a nephrotoxic glutathione conjugate in rats. Toxicol Appl Pharmacol 2000;169:102–113.
68. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry. in vivo drug metabolism/drug interaction studies - study design, data analysis, and recommendations for dosing and labeling. 1999.
69. U.S. Department of Health and Human Services, Food and Drug Administration. Draft guidance for industry: drug interaction studies - study design, data analysis, and implications for dosing and labeling. 2006.
70. Zhang L, Zhang YD, Zhao P, *et al*. Predicting drug–drug interactions: an FDA perspective. AAPS J 2009;11:300–306.
71. Huang SM, Strong JM, Zhang L, *et al*. New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters and the guidance process. J Clin Pharmacol 2008;48:662–670.
72. Dierks EA, Stams KR, Lim HK, *et al*. A method for the simultaneous evaluation of the activities of seven major human drug metabolizing cytochrome P450s using an *in vitro* cocktail of probe substrates and fast gradient liquid chromatography tandem mass spectrometry. Drug Metab Dispos 2001;29:23–29.
73. Turpault S, Brian W, Van Horn R, *et al*. Pharmacokinetic assessment of a five-probe cocktail for CYPs 1A2, 2C9, 2C19, 2D6 and 3A. Br J Clin Pharmacol 2009;68: 928–935.
74. Zhou H, Davis HM. Risk-based strategy for the assessment of pharmacokinetic drug–drug interactions for therapeutic monoclonal antibodies. Drug Discov Today 2009;14:891–898.
75. Wells DA. High throughput bioanalytical sample preparation: methods and automation strategies. Amsterdam: Elsevier; 2003.
76. U.S. Department of Health and Human Services, Food and Drug Administration. General principles of software validation. Final guidance for industry and FDA staff. 2002.
77. U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for industry part 11, electronic records: electronic signatures - scope and application. 2003.
78. Bansal SK, Layloff T, Bush ED, *et al*. Qualification of analytical instruments for use in the pharmaceutical industry: a scientific approach. AAPS PharmSciTech 2004;5. Article 22.
79. Huber L. Validation of computerized analytical systems, Buffalo Grove (IL): Interpharm Press; 1995.
80. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry, investigators, and reviewers - exploratory IND studies. 2006.
81. Buchholz B. Bioanalytics for Human Microdosing In: Muddiman D, Cohen L, editors. Volume 5, Bioanalytical approaches: LC-MS, NMR, isotope methods, spectroscopy, ligand-binding and other analytical chemistry methods. New York: John Wiley and Sons; 2011. pp. 599–626.
82. Garner RC. Less is more: the human microdosing concept. Drug Discov Today 2005; 10:449–451.
83. Balani SK, Nagaraja NV, Qian MG, *et al*. Evaluation of microdosing to assess pharmacokinetic linearity in rats using liquid chromatography-tandem mass spectrometry. Drug Metab Dispos 2005;34:384–388.

84. Ni J, Ouyang H, Aiello M, *et al*. Microdosing assessment to evaluate pharmacokinetics and drug metabolism in rats using liquid chromatography-mass spectrometry. Pharm Res 2008;25:1572–1582.
85. McLean MA, *et al*. Accelerating drug development: methodology to support first-in-man pharmacokinetic studies by the use of drug candidate microdosing. Drug Dev Res 2007;68:14–22.
86. Hoos JS, Sudergat H, Hoelck JP, *et al*. Selective quantitative bioanalysis of proteins in biological fluids by on-line immunoaffinity chromatography-protein digestion-liquid chromatography-mass spectrometry. J Chromatogr B 2006;830:262–269.
87. Andersson L. Molecular imprinting for drug bioanalysis. A review on the application of imprinted polymers to solid-phase extraction and binding assay. J Chromatogr B 2000;739:163–173.
88. Rindgen D, Nakajima M, Wehrli S, *et al*. Covalent modifications to 2′-deoxyguanosine by 4-oxo-2-nonenal, a novel product of lipid peroxidation. Chem Res Toxicol 1999;12:1195–1204.
89. Mirzaei H, Regnier FE. Enhancing electrospray ionization efficiency of peptides by derivatization. Anal Chem 2006;78:4175–4183.
90. Van den Broek I, Sparidans RW, Schellens JHM, *et al*. Quantitative bioanalysis of peptides by liquid chromatography coupled to (tandem) mass spectrometry. J Chromatogr B 2008;872:1–22.
91. Rodriguez-Cabaleiro D, Uytfanghe KV, Stove V, *et al*. Pilot study for the standardization of insulin immunoassays with isotope dilution-liquid chromatography/tandem mass spectrometry. Clin Chem 2007;53:1462–1469.
92. Hu X, Kamberi M. Quantification of apolipoprotein A-I mimetic peptide D-4F in rabbit plasma by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry. J Chromatogr B 2009;877:1946–1950.
93. Alvarez-Salas LM. Nucleic acids as therapeutic agents. Curr Topics Med Chem 2008;8:1379–1404.
94. Lin ZJ, Li W, Dai G. Application of LC-MS for quantitative analysis and metabolite identification of therapeutic oligonucleotides. J Pharm Biomed Anal 2007;44:330–341.
95. Deng P, Chen X, Zhang G, *et al*. Bioanalysis of an oligonucleotide and its metabolites by liquid chromatography-tandem mass spectrometry. J Pharm Biomed Anal 2010;52:571–579.
96. Buscher BAP, Gerritsen H, van Scholl I, *et al*. Quantitative analysis of tenecteplase in rat plasma samples using LC-MS/MS as an alternative to ELISA. J Chromatogr B 2007;852:631–634.
97. Chang D, Kolis SJ, Linderholm KH, *et al*. Bioanalytical method development and validation for a large peptide HIV fusion inhibitor (Enfuvirtide, T-20) and it metabolite in human plasma using LC-MS/MS. J Pharm Biomed Anal 2005;38:487–496.
98. Van den Broek I, Sparidans RW, Schellens JHM, *et al*. Enzymatic digestion as a tool for the LC-MS/MS quantification of large peptides in biological matrices: measurement of chymotryptic fragments from the HIV-1 fusion inhibitor enfuvirtide and its metabolite M-20 in human plasma. J Chromatogr B 2007;854:245–259.
99. Spooner N, Lad R, Barfield M. Dried blood spots as a sample collection technique for the determination of pharmacokinetics in clinical studies: considerations for the validation of a quantitative bioanalytical method. Anal Chem 2009;81:1557–1563.
100. Barfield M, Spooner N, Lad R, *et al*. Application of dried blood spots combined with HPLC-MS/MS for the quantification of acetaminophen in toxicokinetic studies. J Chromatogr B 2008;870:32–37.
101. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry, inspecting out-of-specification (OOS) test results for pharmaceutical production. 2006 Oct.