Review Article

Theme: Best Practices for Bioanalytical Methods: Recommendations from the Global Bioanalysis Consortium
Guest Editors: Binodh DeSilva and Philip Timmerman

Stability: Recommendation for Best Practices and Harmonization from the Global Bioanalysis Consortium Harmonization Team

Nico van de Merbel,^{1,10} Natasha Savoie,² Manish Yadav,³ Yoshiaki Ohtsu,⁴ Joleen White,⁵
Maria Francesca Riccio,⁶ Kelly Dong,⁷ Ronald de Vries,⁸ and Julie Diancin⁹

Received 19 August 2013; accepted 17 January 2014; published online 19 February 2014

Abstract. This paper provides a comprehensive overview of stability-related aspects of quantitative bioanalysis and recommends science-based best practices, covering small and large molecules as well as chromatographic and ligand-binding assays. It addresses general aspects, such as the use of reference values, transferability and treatment of failing stability results, and also focuses on specific types of stability assessment: bench-top, freeze/thaw and long-term frozen stability, stock stability, extract stability, stability in whole blood, tissue and urine, and stability of endogenous analytes, in special matrix types and in incurred samples.

KEY WORDS: GBC; regulated bioanalysis; stability assessment.

INTRODUCTION

It has been long recognized that ensuring analyte stability is of crucial importance in the use of any quantitative bioanalytical method. As analyses are usually not performed directly after collection of the biological samples, but after these have been processed and stored, it is essential that analyte stability be maintained over the relevant storage conditions in order to ensure that the obtained concentration results adequately reflect those directly after sampling. For this reason, stability assessment has always been an important part of bioanalytical method validation. Several key documents, including journal publications (1,2), regulatory guidelines (3–6), and white papers (7–15) have appeared over the past decades that provide guidance to the bioanalytical scientist on how to approach stability assessment in the development and validation of bioanalytical methods. Although the general principles of stability assessment have thus been reasonably well covered, the topic is broad and sometimes complicated, and some specific areas remain in

which there are no or conflicting recommendations or in which a scientific basis is lacking or not described.

In line with the general goal of the Global Bioanalysis Consortium to provide scientific guidance on bioanalytical issues, the aim of the current paper is to provide a comprehensive overview of stability-related aspects of quantitative bioanalysis and to recommend science-based best practices. Since the challenges leading to instability of small and large molecules differ, the proposed approaches may also be different for both molecule classes and/or for the assay type used (chromatographic or ligand-binding). Where applicable, this has been specifically indicated in the text. The recommendations given here are meant for full validations in regulated bioanalysis. Depending on the purpose of the analysis, a more limited validation and a less extensive stability assessment may be applicable.

Next to a comprehensive evaluation in the following sections, Table I summarizes the key recommendations for quick reference.

LEADING PRINCIPLES

Stability is not only related to the chemical integrity of a molecule, but also to other factors that may affect its concentration during sampling and storage, including solvent evaporation, adsorption to containers or collection materials, precipitation, and non-homogeneous distribution over a sample. It is, therefore, not only the absence of analyte degradation that defines sufficient stability, but rather the constancy of the analyte concentration over time in general. Since ligand-binding assays detect a binding event rather than the molecule itself, for this assay type, it is actually the constancy of the immunoreactivity in a sample that

¹ Bioanalytical Laboratory, PRA, Westerbrink 3, 9405 BJ, Assen, The Netherlands.

² Algorithme Pharma, Montreal, Canada.

³ Alkem Laboratories, Mumbai, India.

⁴ Astellas Pharma, Osaka, Japan.

⁵ Biogen Idec, Cambridge, MA, USA.

⁶ CAEP, Sao Paulo, Brazil.

⁷ GlaxoSmithKline, Shanghai, China.

⁸ Janssen Research & Development, Beerse, Belgium.

⁹ PPD, Middleton, WI, USA.

¹⁰ To whom correspondence should be addressed. (e-mail: merbelnicovande@praintl.com)

Table I. Summary of the key recommendations for stability assessment in regulated bioanalysis

Item	Remark*
General	
• Stability assessment should cover all relevant conditions encountered in practice	A
• Storage duration should be at least equal to the maximum storage period for any individual study sample	A
• Deviation of the result for a stored sample from the reference value should not exceed 15% (chromatography) or 20% (binding assays)	S/L
• Two concentration levels (low and high) suffice for stability assessment; stability at an over-curve level is not necessary unless scientifically called for	A
• A single time point suffices for each stability assessment with an appropriate number of replicates	A
• Stability results should generally not be extrapolated to other storage conditions	A
Reference values	
• Results for stored samples can be compared to nominal or $t=0$ values	A
• Fresh calibrators are essential for long-term stability assessment, but calibrators stored frozen suffice for other stability determinations as long as stability under frozen conditions is confirmed	A
Transferability	
• Stability results obtained in other laboratories can be used if storage conditions are similar and if the assessment has been performed in an acceptable way	A
Failing results	
• Stability results should be rejected in the case of an analytical error or failing calibration or QC results and be accepted otherwise	A
• Stability results not meeting the criteria indicate that investigated storage conditions are unsuitable	A
• Possible analytical outliers can be investigated by re-analysis in duplicate	A
Bench-top, freeze/thaw and long-term frozen stability	
• Storage and analysis conditions should mimic the situation for study samples	A
• Frozen stability at a lower temperature is not needed if already demonstrated at a higher temperature, unless scientifically required	S
Stock stability	
• Stability assessment is needed for lowest and highest concentrations which will be stored in practice	A
• Stability assessment is needed for the conditions used for long-term storage and for bench-top use	A
• Deviation of the result for a stored sample from the reference value should not exceed 10%	S
• Stability assessment for internal standards is only required when scientifically called for	S
Extract stability	
• Assessment of relative stability (against stored extracts of calibrators) suffices as long as extracts of calibrators and QCs will be stored together with study samples	S
Whole blood stability	
• Stability assessment is generally not necessary if stability in plasma/serum has been demonstrated under the same conditions, unless the analyte is known to behave differently in the presence of blood cells	A
• Stability assessment can be performed by direct analysis of whole blood or by harvesting and subsequent analysis of plasma/serum	A
Tissue stability	
• Stability cannot be demonstrated for intact tissue, but only for tissue homogenate	A
• Storage of study samples in the form of homogenate is recommended	A
Endogenous analytes	
• Stability assessment should be similar to that for xenobiotics, whenever possible	A
• Stability should be assessed using the authentic matrix and the authentic analyte	A
Deviating matrix types	
• Additional stability assessment should be considered if the biological matrix in a particular study deviates considerably from a normal control matrix and if this deviation in composition is likely to impact analyte stability	A
Incur sample stability	
• Stability in incurred samples should be considered in case of possible differences in stability in spiked and incurred samples	S

QC quality control

*Relevant for small molecules (S), large molecules (L), or all types of molecules (A)

determines stability. This means that the three-dimensional biological integrity of the analyte molecule, which also determines the analytical response, is important to be maintained and that stability may also depend on the reagents used in the assay.

All conditions during sample collection and storage should be such that analyte stability is ensured. Appropriate stabilizing measures, if any (e.g., decreased temperature, protection from light, addition of stabilizers, use of

appropriate collection and storage containers), should be defined during method development. Stability assessment during method validation should only serve to confirm the suitability of the optimized experimental conditions.

Stability can be assessed by subjecting spiked and/or incurred samples to a particular storage condition and subsequently analyzing aliquots of the stored samples against an appropriate reference. The storage conditions to be investigated must be established for each individual case

and should cover all relevant conditions to which study samples will be subjected in practice. These typically include (but are not necessarily limited to): stability in the biological matrix, *e.g.*, bench-top storage, long-term frozen storage and repeated freeze/thaw cycles; stability in sample extracts (where applicable); and stability of stock and derived solutions that are used to prepare calibrators and quality control (QC) samples. There is no need to determine analyte stability at conditions that will not be encountered in standard practice.

The duration of the storage period at a particular condition for which stability is demonstrated, has to cover and thus be at least equal to the maximum period that will be used to store study samples at this condition. Where necessary, during the lifetime of a method, stability assessments can be added or storage periods prolonged to cover expanding needs. Ideally, stability results are available before the first study samples are obtained because running a study without knowing the appropriate storage and stabilizing conditions poses a risk to generate invalid concentration results.

COMMON UNDERSTANDINGS

In order to conclude that an analyte is stable at a particular condition, there should be no discernible difference between the result for a stored sample and the result for a corresponding fresh sample. To evaluate this, different approaches can be followed. By simultaneous replicate analysis of stored and freshly prepared samples and calculating an (*e.g.*, 90%) confidence interval for the difference in response, a statistically significant difference as a result of storage can be assessed (16). However, current practice across the industry is to compare the analytical result for a stored biological sample to the reference value and to allow the difference to be similar to the generally accepted maximum bias for (unstored) QC samples from the nominal value, *i.e.*, generally $\pm 15\%$ for chromatographic and $\pm 20\%$ for ligand-binding assays, as per current guidelines (3–6).

In practice, two concentration levels are generally used to assess stability: a relevant low and a relevant high concentration, which are typically equal to the low and high QC levels that are used to determine method precision and accuracy. Although stability often is not concentration-dependent, there may be situations in which it is advisable to also investigate analyte stability at a relevant over-curve concentration, such as the dilution QC level, in particular, if a majority of the study samples are known or expected to have concentrations exceeding the standard calibration range. This should be considered on a case-to-case basis; examples could include urine assays for relatively hydrophobic analytes which may show concentration-dependent adsorption or precipitation, or methods for large molecules that could be affected by aggregation or prozone effects at high analyte levels.

A stability assessment at a single time point per storage condition per concentration level is considered sufficient, although it should be realized that analysis after multiple time points will provide more detailed information, which can be helpful to interpret the stability profile of an analyte. A sufficient number of replicates should be performed to obtain a reliable average result for any stability assessment. Current

practice appears to be that each stability assessment is typically performed in triplicate as a minimum. For methods with a large intrinsic variability, increasing the number of replicates could be considered to increase the level of confidence in the result and avoid drawing incorrect conclusions.

Stability of a particular compound is determined by temperature, exposure to light, the matrix (including anti-coagulant and the presence of stabilizing additives) and the type and composition of the sample container. Results generally should not be extrapolated to other conditions, although this should be based on scientific judgment. For example, if stability has been demonstrated under lighting conditions, storage in the dark will typically be acceptable as well. Although accelerated stability testing (*e.g.*, by storage at 37°C) is a useful tool in method development to build up knowledge of the stability properties of a compound, it should not be used during method validation to replace stability assessment at the actual storage temperature. The properties of the biological matrix used for stability assessment should properly mimic the situation of study samples. Any major alteration of the biological matrix used for the stability experiments, such as stripping with active carbon, essentially leads to another matrix and is unacceptable. For analytes that are known to be susceptible to enzymatic degradation in the matrix, the use of sufficiently fresh matrix is recommended for stability assessment to avoid overlooking enzyme-mediated degradation, which might be decreased in older lots of the matrix. Scientifically, the effect of the counter ion of the anti-coagulant on analyte stability can generally be considered negligible, at least for small molecules quantified with chromatographic assays, and its change does not require additional stability assessment (15). For large molecules sensitive to pH perturbations, a different counter ion could theoretically induce folding differences and thus affect stability or immunoreactivity.

For multi-analyte assays, it is most practical, although not scientifically required, that all analytes are added to the samples used for stability assessment. If all analytes are added together to the stability samples, it is generally acceptable that they all are present at their respective low or high concentrations. For analytes that could be converted to one another during storage or analysis, it is recommended to evaluate the degree of conversion from one analyte to the other and define the *in vivo* concentration ratio of the analytes for which conversion (if any) is considered acceptable. This can be done by assessing stability with single analytes or with multiple analytes spiked together at a maximum, clinically relevant concentration ratio.

BEST PRACTICES

Reference Values

Analyte stability results are typically obtained by comparing the concentration after storage to a reference value, which can be either the theoretical concentration or the concentration experimentally determined in an aliquot of the sample which has not been subjected to storage ($t=0$ value). Comparison to the $t=0$ value rules out any bias introduced by the spiking process and, therefore, gives a direct estimate of

the effect of storage (although a bias may be introduced by the experimental determination of the $t=0$ value). Comparison to the theoretical value, which is recommended by current regulatory guidelines (3–6), essentially is a combination of the bias caused by storage and pre-analytical bias because of the spiking process. Both approaches are scientifically correct, but for the latter approach, due attention has to be paid to the effect that the bias originating from the spiking process may have on the final result. Therefore, it is recommended that the concentrations of the stability samples be measured prior to the stability experiment to allow evaluation of the accuracy of spiking. In order to avoid incorrect conclusions about stability, it is advisable to consider preparation of a new set of stability samples if the bias found for the spiked samples exceeds a specific value. The magnitude of this value will depend on the risk one accepts, but needs to be more stringent than the usual 15% or 20% acceptable bias (10).

Expressing stability by comparison to the corresponding $t=0$ value can be regarded as equally acceptable. Also in this case, it is advisable to check the $t=0$ concentrations and to consider preparation of a new set of stability samples if their concentrations deviate relatively much from the corresponding nominal values, to avoid that concentrations after storage are outside the normal acceptance criteria. In some situations, comparison to the theoretical value is impossible because the results cannot be directly derived from a calibration curve. Examples include the assessment of the stability of stock and working standard solutions, the stability in whole blood (if performed by preparation and subsequent analysis of plasma), and the stability of endogenous analytes. In these cases, comparison to the corresponding $t=0$ value is the only option.

The use of freshly prepared calibrators for stability assessment, as specified by regulatory guidelines (3–6), is especially important for the determination of (long-term) frozen stability, since the use of calibrators that are stored frozen prior to analysis can lead to a similar analyte instability profile in the stability samples and the calibrators, with the risk that instability passes unnoticed. In this context, a workable definition of “fresh” could be: prepared on the day of the stability assessment, unfrozen, and stored using any necessary stabilizing precautions (such as protected from light or on ice). For other stability assessments, the use of calibrators that are not freshly prepared but stored frozen until their use can be regarded as scientifically appropriate if the stability of the calibrators over the period of frozen storage is ensured.

Transferability of Stability Results

As analyte (in)stability is determined by physicochemical parameters (temperature, time, matrix composition, exposure to light), stability results can be considered universally valid. Therefore, it is scientifically justifiable to refer to stability results obtained elsewhere, if critical storage conditions are similar and the stability assessment has been performed in a scientifically sound and traceable way.

For analytes with a clear stability profile, slight variations in storage conditions (*e.g.*, because of seasonal fluctuations in temperature, humidity, and lighting conditions) will be an

acceptable risk, both within a single laboratory and between different laboratories and do not warrant additional stability investigations. In the case of larger variations in storage conditions between analytical sites or between the analytical site and the sample collection site, it is recommended to either repeat the stability test in each new lab or to cover all different storage conditions in the original stability assessment. Whether or not a stability experiment will need to be repeated should be considered case-by-case and will also depend on the effort it will take to cover the necessary storage conditions during the original validation, compared to simply repeating the experiment.

Any acceptably validated method may be considered to yield reliable results and thus, the acceptability of transfer of stability results does not need identical analytical methods at the different sites. An exception is transferring stability results between methods that use protein–protein interactions (such as ligand-binding assays and immunoprecipitation extractions) and that employ different reagents and/or incubation times, since binding surfaces and kinetics may impact the ability to measure the analyte across methods during stability assessments.

Treatment of Failing Stability Results

In the same way as study samples, stability samples are analyzed in an analytical run containing calibrators and QC samples. If the calibration curve and/or QCs fail the acceptance criteria or if there is a documented error in preparation, then the stability results are invalid and the assessment should be performed again, regardless of whether the stability assessments would have passed. If the run is valid, but the stability results do not meet the pre-defined acceptance criteria, the investigated storage conditions should be considered unsuitable. Where needed, shorter storage periods, lower storage temperatures, and/or the addition of stabilizing agents can be investigated and additional stability testing for the adapted storage conditions performed.

If a failed assessment is suggestive of an analytical outlier, the recommended approach is to investigate this by repeating the experiment in duplicate. Reporting all data and making a scientific judgment of the total data set, based on pre-defined criteria, is appropriate. In some cases, repeating the experiment may be unnecessary for the final decision. One example is a time course where the early and late time points pass the criteria while the middle time point fails. It may be scientifically justifiable to accept the longer time without a repeated assessment depending on the magnitude of the discrepancy and overall data trend.

Bench-Top, Freeze/Thaw and Long-Term Frozen Stability in Biological Matrix

Sufficient analyte stability in the biological matrix during the different phases of storage and analysis is essential to obtain reliable bioanalytical results. The most important phases are the following: (1) storage of the thawed sample at the laboratory bench prior to analysis, (2) (repeated) thawing and refreezing of the sample, and (3) frozen storage between sampling and analysis, which may vary widely in duration from, *e.g.*, less than 1 day to several years. Current

regulatory guidelines include recommendations with different levels of detail on how to determine stability under these conditions, such as bench-top stability for 4 to 24 h and a minimum of three freeze/thaw cycles (3). Scientifically, however, it is more important to cover the maximum storage duration and number of cycles that are relevant for the particular application rather than following prescribed conditions, which may not be encountered in practice. In addition, it is essential to assess stability at the conditions that are used for study samples. For example, if samples are intended to be thawed at 37°C or in the refrigerator rather than at ambient temperature or if thawed samples will be stored on melting ice rather than at the bench-top, these conditions should also be used for stability assessment.

For frozen stability, there has been some debate about the need to cover storage at lower temperatures (such as < -70°C) when stability at a higher temperature (e.g., -20°C) has been demonstrated (8). Molecules will be chemically more stable at a lower storage temperature, based on Arrhenius' law of reaction kinetics, but proteins may undergo (increased) denaturation and partial loss of their three-dimensional structure during freezing at lower temperatures. For small molecules, this could theoretically affect analyte extractability from a protein-containing matrix. Since analyte-protein binding for small molecules will typically be rigorously broken during sample pretreatment, it is reasonable to assume that stability at lower temperatures generally will not differ from stability at higher temperatures. Assessment of stability at a lower storage temperature will, therefore, typically not be necessary, except when scientifically called for. The possible denaturation of large molecules could influence their recognition by ligand-binding assays. This denaturation is most likely to take place during the actual freezing and thawing processes rather than during frozen storage at a constant temperature. Therefore, the assessment of the effect of repeated freeze/thaw cycles at different temperatures may be more important than the assessment of long-term frozen storage stability.

Stock Stability

Since study samples are quantified against spiked calibrators, the reliability of bioanalytical results, in general, not only depends on sufficient analyte stability in the biological matrix, but also on the stability of the solutions used for spiking. Generally, a primary stock solution is prepared for each analyte in an aqueous solution, in an organic solvent, or in a solvent mixture, which can subsequently be diluted to one or a series of working standards, all of which may be used for spiking. To justify the use of a stock or diluted solution after storage, analyte stability in these solutions should be demonstrated for the appropriate period and at the actual storage conditions, not only because of potential analyte degradation but also because repeated, long-term use of a solution may lead to concentration changes due to solvent evaporation. If solutions at different analyte concentrations are stored, it is recommended to perform stability assessments at the lowest and highest concentrations encountered in practice, as adsorption effects may be concentration-dependent. In order to cover the different storage conditions to which stock and derived solutions are subjected, it is

necessary to investigate stability at the conditions both for long-term storage (typically refrigerated or frozen) and for daily use at the laboratory bench (typically at ambient conditions).

In contrast to the stability in a spiked biological matrix, stock stability for small molecules is not assessed by reference to a calibration curve but by comparing the analytical response of a stored solution to that of a freshly prepared reference solution. The use of an existing stock solution as the reference, even when within its proven stability period, increases the risk of misinterpreting the stability of a stored stock solution and should be avoided (8). To limit error propagation in the preparation process of calibrators and QCs, it is recommended that the acceptable difference between the responses of fresh and stored solutions should be tighter than for spiked biological samples and not exceed $\pm 10\%$, as was recently described (5). Similarly, in order to be able to reliably demonstrate such a small difference between fresh and stored solutions, the coefficient of variation for each set of replicates should be limited and preferably no more than 10%. With regard to internal standards, which are often applied in chromatographic methods, very frequently, a stable-isotope labeled form of the analyte or a close chemical analogue is used and the stability properties of these molecules typically are very similar to those of the analyte, when stored at identical conditions. In addition, internal standards are used for response normalization in a single run only and the suitability of the internal standard solution is thus evaluated daily by reviewing the results of the bioanalytical runs. Therefore, the continued long-term storage stability of internal standard solutions is much less critical than for standards solutions of the analytes. Often, it may not be necessary to separately assess the stability of internal standard solutions, if analyte stability has been demonstrated and, for stable-isotope labeled forms, if no isotope exchange occurs during storage. Internal standard stability assessment may be required in individual cases, when scientifically called for, such as in the case of molecules, solvents and/or storage conditions that are very different from those of the analyte. In such a case, wider acceptance criteria than for analyte stock stability are fully acceptable.

For reference material received as a solution, the stability information provided by the manufacturer is acceptable for use and does not need to be regenerated at the analytical lab, provided that the manufacturer's storage conditions are followed. When possible, dilutions should be prepared in the same solvent or buffer as the provided stock solution to build on the manufacturer's stability documentation. If the manufacturer's stability information covers a range of concentrations and temperatures that include the diluted stock concentration and storage, stability assessment does not need to be repeated.

Extract Stability

For methods using some form of sample extraction during analysis, such as most chromatographic assays, the assessment of analyte stability in the extracts is needed to ensure complete coverage of the relevant storage conditions. Post-preparative stability or extract stability refers to the stability of the analyte in sample extracts in general, while on-

instrument stability refers to analyte stability in sample extracts during storage in an autosampler or any other instrument for serial analysis, prior to their actual analysis. Absolute (post-preparative or on-instrument) stability is defined as the absolute constancy of a concentration over time, as measured against a freshly extracted calibration curve, while relative (post-preparative or on-instrument) stability is defined as the constancy of a concentration over time relative to a calibration curve that is extracted and stored together with the sample extracts.

Post-preparative stability needs to be determined for all conditions to which extracts will be subjected. The stability of extracts during storage prior to placement in the autosampler, such as a few hours at room temperature, does not need to be assessed as a standard. As long as extracted study samples are always stored together with and under identical conditions as the extracted calibration and quality control samples, it is not necessary to establish the absolute post-preparative or on-instrument stability, because any degradation over time will occur to the same extent for study samples, calibrators, and QCs. In such a case, it will suffice to assess relative post-preparative or on-instrument stability, as long as sufficient sensitivity is maintained for quantification.

Re-injection reproducibility refers to the possibility to re-inject a complete or partial batch of samples into the analytical instrument. This is typically done by analyzing a set of samples, re-injecting them together with the calibrators, and calculating the result against the re-injected calibration curve and/or against a freshly extracted curve. It should be noted that this is, in fact, a reproducibility assessment rather than a stability assessment.

Whole Blood Stability

Although analytes being measured in plasma or serum will be stored in whole blood for some time during the sampling process and information about their stability in this period of time would be needed, there is accumulating evidence that there is little difference between stability in plasma/serum and stability in whole blood except for some well-defined classes of (small) molecules such as *N*-oxides and hydroxamic acids (12). Whole blood stability testing may therefore not be strictly necessary in all situations as long as plasma/serum stability under the same storage conditions has been demonstrated. It is recommended that stability in whole blood should be assessed at least for compounds belonging to these classes and for compounds with borderline stability in plasma/serum.

Two approaches for whole blood stability testing exist as follows: (1) the analysis of plasma/serum after removal of cellular material and (2) the analysis of whole blood itself. Both approaches have their own advantages and disadvantages (11,12). The analysis of plasma/serum best mimics the actual sampling procedure and gives combined information about the stability and the *ex vivo* redistribution of analytes over plasma and blood cells during the sample collection procedure. In addition, it allows the use of the validated plasma or serum method. When using this approach, it is essential that the analyte distribution between blood cells and plasma should be at equilibrium at $t=0$ in order to prevent the stability results from being affected by an ongoing

distribution process during the storage period. Therefore, for analytes with significant partitioning into the cellular fraction, spiked whole blood samples should be incubated for a sufficient period to reach this equilibrium prior to executing the stability test. The direct analysis of whole blood gives information about stability only, allows the use of nominal concentrations as the reference, and does not suffer from any distribution phenomena. However, it might need a different analytical approach and sometimes additional method development. Also, samples cannot be analyzed using the validated plasma/serum methodology.

From a scientific point of view, both approaches are suitable to assess stability in whole blood. If the analysis of plasma or serum gives acceptable results, stability in the original whole blood sample can be considered to have been demonstrated. Where necessary, both approaches can be combined to obtain more detailed information. Whichever approach is followed, care should always be taken to prevent hemolysis during spiking and incubation. Therefore, the volumes of organic solvents for spiking the whole blood samples should be kept to a minimum. In addition, in order to properly mimic the situation during sampling, it is recommended that spiked whole blood should be warmed to 37°C prior to stability testing.

Tissue Stability

Next to the general considerations for stability assessment, tissue samples have a number of specific issues that need to be taken into account. As intact tissues cannot be homogeneously spiked, stability assessment has to be performed by preparing blank tissue homogenate, spiking it with the analyte and subsequently subjecting it to the relevant stability experiments. The composition and volume of the homogenization solvent can influence the resulting analyte stability profile and will need to be optimized in the stage of method development. Since it may vary from tissue to tissue and from species to species, stability needs to be established for each individual new matrix. Because of the inherent absence of stability data in the intact tissue, it is advisable to homogenize tissues from dosed animals as soon as possible after collection and store the samples as homogenate, in order to cover as much of the storage period as possible by actual stability results. There might be logistical limitations that complicate the homogenization of tissue samples immediately after collection. If this is the case, then it is advised to store the intact tissues at $< -70^{\circ}\text{C}$ rather than at -20°C until homogenization and to still try to restrict the storage period from sampling until homogenization to a minimum.

Urine Stability

Since the native pH of urine can vary between approximately 4 and 8.5, it is important that stability assessments take this into account. It is recommended that stability of the analyte in urine be assessed at three pH values, *e.g.*, pH 4, 6.5, and 8.5, to properly evaluate the stability profile under actual physiological conditions. Since adsorption problems in urine are likely to be more pronounced than in more protein-rich matrices, it is especially important that the same container material and additives, if any, be used for the stability

assessments and the actual urine sampling. The experiments can be conducted in smaller containers (*e.g.*, 30 mL or even less) while sampling is done in larger containers (often 1 or 2 L), as long as the container material and the typical urine volume to container surface area ratio are similar. Since a small urine aliquot is usually transferred to a secondary container which is subsequently sent to the bioanalytical lab, stability should also be tested in the secondary container, if this is made of a different material.

While urine is typically clear and homogeneous at the moment of voiding, considerable amounts of sediment may be formed upon storage, depending on the composition of the urine and the storage temperature. This phenomenon essentially makes urine a non-homogeneous matrix and can have a pronounced effect on the concentrations after storage, particularly when an analyte binds to the sediment and a non-representative sub-sample is taken for analysis. Proper attention should therefore be paid to the optimization of the procedure for urine collection and the taking of representative sub-samples, to ensure that the reported analyte concentration in urine does not deviate from the concentration directly after sampling. Another bioanalytical aspect in which urine may be different from other matrices, at least for small organic molecules, is the possible presence of higher amounts of phase-2 metabolites that can convert back to the analyte during storage and thus give rise to overestimation of analyte concentrations. To address this, additional stability experiments using samples spiked with relevant concentrations of these metabolites or using incurred samples should be considered, where it should be taken into account that stability of phase-2 metabolites can be pH-dependent.

Stability of Endogenous Analytes

The stability assessment of endogenous analytes (*e.g.*, biomarkers, endogenous therapeutics) has some special features because of the typical absence of analyte-free biological matrix. Still, it should follow the general approach, as described above, whenever possible (17). It is essential that the samples used for stability determination should be the authentic biological matrix containing the authentic analyte, although it is recognized that calibration samples generally are in a different form (surrogate matrix or surrogate analyte), because of the usual presence of analyte in the authentic matrix. A typical approach would be to screen a number of authentic matrix lots for low analyte levels and select a sample to be used as the low concentration for stability assessment. The low-concentration stability sample could then be spiked to prepare the high-concentration stability sample or, alternatively, an incurred sample with a suitable high concentration could be selected after screening a number of lots of the matrix. Stability will typically have to be expressed against a $t=0$ value because the exact nominal concentration generally is unknown. A complicating factor is that a truly fresh $t=0$ reference sample cannot easily be obtained because of the need for screening, which implies that the samples selected for stability assessment will have been stored for some time and may also have been subjected to one or more freeze/thaw cycles prior to their actual use. If it is the intention not to prepare fresh calibration standards in each run for bioanalysis but to use stored calibrators, stability

should also be assessed in the calibration matrix, if this is different from the matrix of the study samples.

Stability in Deviating Matrix Types

Normally, a biological control matrix, which is used for stability assessment, is obtained from healthy, undosed subjects. A deviating type of a biological matrix can be defined as a form of the matrix, which has a composition that is different from that of the regular control matrix. The investigation of stability in such a deviating matrix should be considered if there is a scientific reason to do so. In practice, this will be the case if the composition of the matrix for a particular study significantly deviates from that of regular control matrix and if it is reasonable to assume that this deviation will have an impact on the stability profile of the analyte of interest. Whether or not this is the case will depend on the characteristics of the analyte and the matrix and needs to be determined for each situation.

In general, cases needing additional stability assessment will be rare, but examples could include hemolytic plasma, matrices from patient populations and matrices containing co-administered drugs (13), or excipients originating from the drug formulation. Situations that could trigger the investigation of stability in deviating matrix types include analyte instability upon contact with hemoglobin or other constituents of hemolyzed blood cells, substantial differences in enzyme activity between healthy and diseased matrices, and a significant impact of a co-administered drug or excipient on sample properties.

Stability in Incurred Samples

The main difference between spiked and incurred samples with regard to stability is the potential presence of unstable metabolites in incurred samples that could convert back to the analyte during sampling, storage and/or analysis, and result in overestimation of analyte concentrations (*e.g.*, *N*-oxides, glucuronides, interconversion of stereo-isomers). This effect is typically not expected for large molecules, but depending on their molecular properties, it is a possibility for small organic analytes. This effect can only be investigated during stability assessment with spiked samples if reference standards of the unstable metabolites are available and added at relevant concentrations to the stability samples. Alternatively, assessment of stability in incurred samples represents a way to evaluate the impact where it should be noted that for methods in plasma or serum, stability in incurred whole blood samples might be relevant as well.

Incurred sample stability (ISS) should not be considered as a goal in itself. The main stability dataset will be stability assessments done using spiked samples but ISS could serve to bridge a possible gap between spiked and incurred samples, when deemed necessary based on the physicochemical and/or metabolic properties of the analyte. As the metabolite profile can be different for different species, it is recommended that the experiments should be conducted in all relevant species, whenever feasible. It should not be required to repeat experiments,

e.g., for different subject populations and disease states, unless there are indications that the abundance of potentially unstable metabolites is significantly increased. Should there be incurred sample reproducibility (ISR) failures in the conducted bioanalyses, then, assessment of ISS could also be used as an investigation strategy for ISR failures, as these might be related to incurred sample instability. Timing of the experiments should be flexible. The ISS experiments can be conducted as soon as the study samples from the first study are available, but can also be conducted in a later stage of drug development, e.g., in phase 2 or 3. Should the experiments demonstrate that there are issues with the method when applied to incurred samples, then the impact of the issues with the method on the bioanalytical data generated using the method should be assessed and documented.

UNRESOLVED ISSUES

No final conclusion was reached about the need for stability assessment during transport of samples. Biological samples are typically transported frozen on dry ice for a limited period of time from, e.g., clinic to laboratory, and at a temperature not exceeding the temperature at which long-term storage stability is assessed. In many cases, the transport conditions will be sufficiently covered by other stability results and there will be no scientific need for separate stability assessment. There may be special situations, however, that call for a separate assessment of analyte stability during the transportation process.

CONCLUSIONS

The purpose of this paper is to provide science-based guidance for stability assessment for full validation in regulated quantitative bioanalysis. First and foremost, it is important that the bioanalytical scientists involved fully understand the properties of the analytes and the biological matrices as well as the impact of all sampling, shipment, storage and analysis conditions, so that appropriate measures can be defined that ensure stability and allow the unbiased determination of analyte concentrations. Although the recommendations given here aim at providing a comprehensive overview of stability-related issues in bioanalysis, additional experiments may be necessary, based on the specific characteristics of an analyte, the special nature of a particular biological matrix or uncommon situations during sample collection and storage. In the end, the required conditions for sampling and storage as well as the necessary validation experiments to demonstrate that stability is uncompromised at these conditions should always be based on a sound scientific judgment and not merely on a standard protocol.

ACKNOWLEDGMENTS

The authors want to express their gratitude to Marian Kelley, Lauren Stevenson, and Binodh DeSilva for their valuable discussions during the preparation of the manuscript.

REFERENCES

1. Shah VP, Midha KK, Dighe S, *et al.* Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Conference report. Eur J Drug Metab Pharmacokinet. 1991;16:249–55.
2. Shah VP, Midha KK, Findlay JW, *et al.* Bioanalytical method validation—a revisit with a decade of progress. Pharm Res. 2000;17:1551–7.
3. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine (2001) Guidance for industry: bioanalytical method validation
4. European Medicines Agency, Committee for Medicinal Products for Human Use (2011) Guideline on bioanalytical method validation
5. Agência Nacional de Vigilância Sanitária, Brazil. Resolução – RDC No 27. May 2012.
6. Ministry of Health, Labour and Welfare, Japan (2013) Guideline on bioanalytical method validation in pharmaceutical development
7. 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;20:1885–900.
8. 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:E30–42.
9. Bansal S, DeStefano A. Key elements of bioanalytical method validation for small molecules. AAPS J. 2007;9:E109–14.
10. 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:E117–22.
11. Lowes S, Jersey J, Shoup R, *et al.* Recommendations on: internal standard criteria, stability, incurred sample reanalysis and recent 483s by the Global CRO Council for bioanalysis. Bioanalysis. 2011;3:1323–32.
12. Freisleben A, Brudny-Klöppel M, Mulder H, *et al.* Blood stability testing: European Bioanalysis Forum view on current challenges for regulated bioanalysis. Bioanalysis. 2011;3:1333–6.
13. Lowes S, Boterman M, Doig M, *et al.* Recommendations on bioanalytical method stability implications of co-administered and co-formulated drugs by Global CRO Council for bioanalysis (GCC). Bioanalysis. 2012;4:2117–26.
14. Van Amsterdam P, Companjen A, Brudny-Kloppel M, *et al.* The European Bioanalysis Forum community's evaluation, interpretation and implementation of the European Medicines Agency guideline on Bioanalytical Method Validation. Bioanalysis. 2013;5:645–59.
15. Sennbro CJ, Knutsson M, Timmerman P, *et al.* Anticoagulant counter ion impact on bioanalytical LC-MS/MS assay performance: additional validation required? Bioanalysis. 2011;3:2389–91.
16. Timm U, Wall M, Dell D. A new approach for dealing with the stability of drugs in biological fluids. J Pharm Sci. 1985;74:972–7.
17. Houghton R, Horro Pita C, Ward I, Macarthur R. Generic approach to validation of small-molecule LC-MS/MS biomarker assays. Bioanalysis. 2009;1:1365–74.