Scientific Working Group for Forensic Toxicology (SWGTOX) Standard Practices for Method Validation in Forensic Toxicology

This standard has been adopted by the Scientific Working Group for Forensic Toxicology (SWGTOX) and is intended to reflect a minimum standard of practice. Laboratories choosing to meet this practice must decide on an implementation plan that is conducive to the operation, resources and means of the laboratory.

1. Introduction

This document delineates minimum standards of practice for validating analytical methods in forensic toxicology.[1] Validation is the process of performing a set of experiments that reliably estimates the efficacy and reliability of an analytical method or modification to a previously validated method. The goal of validation is to establish objective evidence that demonstrates a method is capable of successfully performing at the level of its intended use and to identify the method's limitations under normal operating conditions. The steps provided in this document will ensure that the minimum standards of practice for validating analytical methods in forensic toxicology have been performed. While it is acknowledged that method performance may vary somewhat during day-to-day analysis of actual case samples, the validation parameters evaluated with this practice serve as estimates of a method's true performance.

2. Definitions

2.1. *Bias* – the closeness of agreement between the mean of the results of measurements of a measurand and the true (or accepted true) value of a measurand. It is reported as a percent difference. The terms *accuracy* or *trueness* may also be used to describe bias.

2.2. *Blank Matrix Sample* – a biological fluid or tissue (or synthetic substitute) without target analyte or internal standard.

2.2.1. *Fortified Matrix Sample* – a blank matrix sample spiked with target analyte and/or internal standard using reference materials.

2.3. *Calibration Model* – the mathematical model that demonstrates the relationship between the concentration of analyte and the corresponding instrument response.

2.3.1. *Working Range* – the concentration range of that can be adequately determined by an instrument, where the instrument provides a useful signal that can be related to the concentration of the analyte.

[1] SWGTOX will produce a separate document to address validation of methods used in breath alcohol testing.

2.4. *Carryover* – the appearance of unintended analyte signal in subsequent samples after the analysis of a positive sample.

2.5. *Decision Point* – an administratively defined cutoff or concentration that is at or above the method's limit of detection or limit of quantitation and is used to discriminate between positive and negative results.

2.6. *Dilution Integrity* – the assurance that bias and precision are not significantly impacted when a sample is diluted.

2.7. *Fluids* – any liquid biological specimen that is typically pipetted for analysis (e.g., blood, urine, bile, serum, vitreous humor, oral fluid).

2.8. *Interferences* – non-targeted analytes (i.e., matrix components, other drugs and metabolites, internal standard, impurities) which may impact the ability to detect, identify, or quantitate a targeted analyte.

2.9. *Ionization Suppression/Enhancement* – direct or indirect alteration or interference in the instrument response due to the presence of co-eluting compounds.

2.10. *Limit of Detection* – an estimate of the lowest concentration of an analyte in a sample that can be reliably differentiated from blank matrix and identified by the analytical method.

2.11. *Limit of Quantitation* – an estimate of the lowest concentration of an analyte in a sample that can be reliably measured with acceptable bias and precision.

2.12. *Precision* – the measure of the closeness of agreement between a series of measurements obtained from multiple samplings of the same homogenous sample; it is expressed numerically as imprecision.

2.13. *Reference Material* – material, sufficiently homogenous and stable with respect to one or more specified properties, which has been established to be fit for intended use in a measurement process.

2.14. *Stability* – an analyte's resistance to chemical change in a matrix under specific conditions for given time intervals.

2.15. *Tissues* – any solid biological specimen that is generally weighed for analysis (e.g., brain, liver, muscle, hair, meconium).

3. When to Validate Methods

Methods shall be validated when it is necessary to verify a method's performance parameters are fit for use for a particular analysis. Common examples include:

- New analytical method
- Modifications of an established analytical method to improve performance or extend its use beyond that for which it was originally validated (e.g., addition of new compounds to the method's scope)
- To demonstrate equivalence between an established method/instrument and a new method/instrument
- Existing analytical methods that do not currently meet the requirements of this document

The parameters to be evaluated for validation of methods will depend upon the circumstances in which the method is to be used. Likewise, it is recognized that after validation has occurred, methods may be revised. The extent and frequency of revalidation of previously validated methods will depend upon the nature of the intended changes or laboratory policy. See Section 9 for further guidance on revalidation of previously validated methods.

4. Method Development and Optimization

For purposes of this document, method development will be considered in two phases: 1) instrumental and data acquisition/processing parameters and 2) sample preparation. It is essential that validation is conducted with the same analytical conditions and techniques as the final method.
The principles of good laboratory practice and record keeping should be applied to the concepts of this document. This includes documentation of parameters that were evaluated during method development but did not provide acceptable results.

4.1. Development and Optimization of Instrumental and Data Processing Parameters

Instrumental and data processing parameters are defined and optimized through analysis of reference materials of the analyte(s) of interest to achieve the required performance of the instrument.

4.2. Development and Optimization of Sample Preparation Techniques

The sample preparation technique shall be evaluated and optimized using reference materials of the analyte(s) of interest. The primary goal is to demonstrate that the sample preparation steps allow for adequate extraction, detection, identification, and/or quantitation of the analyte(s). Sample preparation shall be evaluated with fortified matrix samples.

5. Establishing a Validation Plan

The laboratory is responsible for ensuring its methods are adequately validated. A validation plan shall be in place prior to starting any validation experiments. The validation plan is separate from a laboratory's standard operating procedure for method validation. The plan shall include the instrumental method(s) and sample preparation technique(s) to be used for a specific method. Further, it shall document the validation requirements of the method, as well as the limits of the method that will allow it to be fit for use. The validation plan provides direction for the experiments that will be performed and acceptance criteria for each parameter. Appendices A and B provide examples of validation plans.

6. Required Validation Parameters Based on Scope of the Method

The scope of forensic toxicology methods is typically categorized as screening methods, qualitative confirmation/identification methods, or quantitative methods. As such, the following validation parameters shall be evaluated:

6.1. Screening (Immunoassay-based):

- Limit of detection
- Precision (at the decision point)
- Dilution integrity (if applicable)
- Stability (if applicable)

6.2. Screening (All others):

- Interference studies
- Limit of detection
- Dilution integrity (if applicable)
- Stability (if applicable)

6.3. Qualitative confirmation/identification:

- Carryover
- Interference studies
- Ionization suppression/enhancement (for applicable techniques, such as LC/MS)
- Limit of detection
- Dilution integrity (if applicable)
- Stability (if applicable)

6.4. Quantitative analysis:

- Bias
- Calibration model
- Carryover
- Interference studies
- Ionization suppression/enhancement (for applicable techniques, such as LC/MS)
- Limit of detection
- Limit of quantitation
- Precision
- Dilution integrity (if applicable)
- Stability (if applicable)

7. Specific Requirements for Conducting Method Validation Experiments

All validation experiments shall be conducted using fortified samples of the matrix for which the method is intended, unless otherwise noted. In some instances (e.g., immunoassay screens), it may be more appropriate to analyze previously characterized human samples instead of fortified samples for selected method validation studies.

Validation studies shall be conducted in a manner similar to casework. This may include conducting validation studies on different days, by different analysts, and ensuring that instruments meet the same daily performance requirements as for casework.

Whenever possible, fortified matrix samples shall be prepared from reference materials that are from a different source (e.g., supplier or lot number) than used to prepare calibration samples. In instances where the same source must be utilized, separate weighings or solutions must be used to prepare these samples.

The following requirements are the minimum for assessing the listed validation parameters in forensic toxicology methods. They are listed alphabetically and not necessarily in procedural order. Some of the validation experiments are demonstrated in Appendices A and B. Section 11 provides guidance on how to efficiently perform validation experiments.

7.1. Bias and Precision

7.1.1. Bias

Bias studies must be carried out for all quantitative methods. These can be conducted concurrently with precision studies. Bias shall be measured in pooled fortified matrix samples using a minimum of three separate samples per concentration at three different concentration pools (low, medium, and high[2]) over five different runs. The bias shall be calculated for each concentration using the following formula:

$$Bias\ (\%)\ at\ Concentration_x = \left[\frac{Grand\ Mean\ of\ Calculated\ Concentration_x - Nominal\ Concentration_x}{Nominal\ Concentration_x}\right] \times 100$$

The maximum acceptable bias is ±20% at each concentration. For some analyses, where less bias is required (e.g., ethanol), a bias of ±10% or better is expected. It is recommended that the same data used in bias studies also be used for precision calculations.

7.1.2. Precision

Precision studies must be carried out for all quantitative methods, as well as at the decision point for immunoassays. These studies

[2] For purposes of this document, low concentrations shall be approximately 3 times the lowest end of the working range of the method and high concentrations shall be within approximately 80% (or more) of the highest end of the working range of the method, unless otherwise noted. Medium concentrations shall be near the midpoint of the low and high concentrations.

can be carried out concurrently with bias studies, if required in the validation plan.

Precision is expressed as the coefficient of variation (% CV). The mean and standard deviation (*s)* of the response is calculated for each concentration to determine the % CV.

$$\%CV = \frac{s}{mean\ response}$$

7.1.2.1. Precision of Immunoassays at Decision Point

For immunoassays that cross-react with a broad class of compounds (e.g., benzodiazepines, opiates, amphetamines), if a laboratory declares to its customers that it is able to detect analytes with low cross-reactivity (less than or equal to the target analyte) using the immunoassay, it is essential to verify its ability to detect these compounds. For example, a benzodiazepine immunoassay targeted for oxazepam typically has low cross-reactivities to other benzodiazepines such as lorazepam. If the laboratory uses this immunoassay kit to screen for lorazepam, it is required to evaluate the assay's ability to reliably detect lorazepam in addition to oxazepam. In contrast, if alprazolam has greater cross-reactivity than oxazepam, there is no requirement to evaluate the ability to detect alprazolam *provided that the decision concentration for alprazolam is not lower than the decision concentration for oxazepam.* This evaluation may require an adjustment or reevaluation of the decision point or the target compound depending on the needs and mission of the laboratory.

At a minimum, precision at the decision point shall be assessed using three separate samples per concentration at three different concentration pools:

- generally no more than 50% below decision point,
- at decision point, and
- generally no more than 50% above decision point over five different runs.

It is understood that immunoassays are matrix-dependent and the concentrations around the decision point should be tighter for a matrix such as urine when compared to others.

The % CV shall not exceed 20% at each concentration. Further, the mean plus or minus two standard deviations for each concentration must not overlap for the decision point to be valid.

It should be noted that the data obtained from these studies are also used to estimate the LOD for immunoassays.

7.1.2.2. Precision of Quantitative Procedures

For quantitative procedures, two different types of precision studies shall be assessed during method validation: within-run precision and between-run precision. At a minimum, precision shall be assessed using three different samples per concentration at three different concentration pools (low, medium, and high) over five different runs.

The % CV shall not exceed 20% at each concentration. It is noted that certain analytical methods (e.g., blood alcohol analysis) may require a much lower coefficient of variation (≤10% CV).

7.1.2.2.1. Within-Run Precision Calculations

Within-run precision is calculated for each concentration separately for each of the five runs. Within-run precision may be calculated using the data from each run's triplicate analyses at each concentration as:

$$Within-run\ CV(\%) = \frac{std\ deviation\ of\ a\ single\ run\ of\ samples}{mean\ calculated\ value\ of\ a\ single\ run\ of\ samples} x100$$

The largest calculated within-run % CV for each concentration will be used to assess within-run precision acceptability.

7.1.2.2.2. Between-Run Precision Calculations

Between-run precision is calculated for each concentration over the five runs. This may be done by using the combined data from all replicates of each concentration as:

$$Between - Run\ CV(\%) = \frac{std\ dev\ of\ grand\ mean\ for\ each\ concentration}{grand\ mean\ for\ each\ concentration} x100$$

7.1.2.2.3. One-Way Analysis of Variation (ANOVA) Approach to Calculate Combined Within-Run and Between-Run Precision

Both within-run and between-run precisions may be calculated using the one-way ANOVA approach with the varied factor (run number) as the grouping variable. The ANOVA calculations can be easily performed using a spreadsheet or a statistical software program.

Using this approach, within-run precision is calculated for each concentration as:

$$Within - run\ CV(\%) = \left[\frac{\sqrt{MS_{wg}}}{grand\ mean\ for\ each\ concentration}\right] x100$$

where MS_{wg} is the mean square within groups obtained from the ANOVA table.

Likewise, between-run precision is calculated as:

$$Between - Run\ CV(\%) = \left[\frac{\sqrt{\frac{MS_{bg} + (n-1) * MS_{wg}}{n}}}{grand\ mean\ for\ each\ concentration}\right] x100$$

where MS_{bg} is the mean square between groups obtained from the ANOVA table and *n* is the number of observations in each group (e.g., *n*=3 if doing triplicate analyses). Appendix A provides an example

of how the ANOVA approach may be used to calculate within-run and between-run precision.

7.2. Calibration Model

The calibration model shall be determined for all quantitative methods. This is accomplished by first determining the range of analyte concentrations over which the method shall be used, sometimes called the *working range*. Within this range, there will be a correlation between signal response (e.g., peak area ratio of analyte and internal standard) and analyte concentration in the sample. The calibration model is the mathematical model that describes this correlation. The choice of an appropriate model (i.e., linear or quadratic) is necessary for accurate and reliable quantitative results.

Calibrator samples are analyzed to establish the calibration model. The use of matrix-matched calibrator samples is encouraged but not required. Regardless of the matrix used to prepare calibrator samples, a laboratory must demonstrate acceptable bias and precision with control samples prepared in all matrices intended to be analyzed by the method (see Section 7.1). For example, blood alcohol methods may demonstrate acceptable bias and precision in whole blood controls using aqueous calibrator samples. Likewise, acceptable bias and precision may be demonstrated using calibrator samples prepared in whole blood but used to quantitate analytes in different matrices (e.g., postmortem tissues, serum, urine).

The calibrator samples shall span the range of concentrations expected. At least six different non-zero concentrations shall be used to establish the calibration model. The concentrations shall be appropriately spaced across the calibration range to establish the most appropriate calibration model. A minimum of five replicates per concentration is required. The replicates to establish the calibration model shall be in separate runs. All data points from the five runs shall be plotted together (using a statistical software package) to establish the calibration model. The origin shall not be included as a calibration point.

The most often used calibration model is the simple linear regression model using the least squares method. However, this model is only applicable when there is constant variance over the entire concentration range. When there is a notable difference between variances at the lowest and highest concentrations, a weighted least squares model or other

appropriate non-linear model shall be applied.[3] This is generally the case when the concentration range exceeds one order of magnitude. Ultimately, the best approach is to use the simplest calibration model that best fits the concentration-response relationship.

Although it has become widespread practice, it is emphasized that a calibration model cannot be evaluated simply via its correlation coefficient (*r*). Instead, a calibration model shall be visually evaluated using standardized residual plots. These allow one to check for outliers that must be eliminated if found to be statistically significant (e.g., outside ±3 standard deviations). Further, residual plots allow one to determine if the variances appear to be equal across the calibration range with a similar degree of scatter at each concentration. They also give an indication if the chosen model adequately fits the data. For example, random distribution of individual residuals around the zero line (homoscedasticity) suggests that a linear model is appropriate.

Finally, there are other appropriate alternatives to evaluate calibration models (i.e., ANOVA lack-of-fit test for unweighted linear models, checking for significance of the second order term in quadratic models, assessment of coefficient of determination for linear models).

If a linear calibration model has been established, fewer calibration samples (i.e., fewer levels or single/fewer replicates) may be used for routine analysis. However, if fewer calibration samples are chosen, the same calibrators (e.g., number, replicates, and concentration level) must be used for the bias and precision studies carried out during the validation studies. Further, they must include the lowest and highest calibration levels used to establish the model, as well as include no fewer than four non-zero calibration points.

Additionally, once the calibration model is established for a validated method, it shall not be arbitrarily changed to achieve acceptable results during a given analytical run. For example, one shall not switch from an unweighted linear model to a weighted linear model in order to adjust for changes in instrument performance.

[3] In general, non-linear models may require additional calibrators to accurately characterize the curve.

7.3. Carryover

Analyte carryover into a subsequent sample may lead to an inaccurate qualitative or quantitative result when using instrumental methods. Carryover must be evaluated during method validation intended for confirmation and/or quantitation, unless a laboratory is constantly addressing carryover in their QA/QC practices.

To evaluate carryover as part of method validation, blank matrix samples are analyzed immediately after a high concentration sample or reference material. The highest analyte concentration at which no analyte carryover is observed (above the method's LOD) in the blank matrix sample is determined to be the concentration at which the method is free from carryover. This concentration shall be confirmed using triplicate analyses.

It is acceptable to limit the carryover study to the highest point of your calibration curve, but even higher concentrations are preferred. If possible, the analytical procedure will be modified to remove any carryover. In cases when it is not possible to eliminate the carryover, the SOP must address how carryover will be managed (e.g., the signal in case samples must be 10 times greater than the signal in a blank sample immediately preceding the case sample or the case samples will be re-extracted and reanalyzed).

7.4. Interference Studies

Interfering substances from common sources must be evaluated in all screening (except immunoassays), qualitative identification, and quantitative methods.

7.4.1. Evaluating Matrix Interferences

Whenever possible, blank matrix samples from a minimum of 10 different sources without the addition of an internal standard (when used in the method) shall be analyzed to demonstrate the absence of common interferences from the matrix. While this approach may detect the more common matrix interferences, it is recognized that less common interferences may not be detected.

7.4.2. Evaluating Interferences from Stable-Isotope Internal Standards

For methods employing stable isotope internal standards, the isotopically-labeled compounds may contain the non-labeled

compound as an impurity. Additionally, the mass spectra of the labeled analogues may contain fragment ions with the same mass-to-charge ratios as the significant ions of the target analyte. In both instances, analyte identification or quantitation could be impacted.

Stable-isotope internal standard interferences shall be assessed by analyzing a blank matrix sample fortified with the internal standard and monitoring the signal of the analyte(s) of interest. Interferences below the LOD of the assay may be insignificant depending on the laboratory's mission.

Likewise, a blank matrix sample fortified with the analyte(s) at the upper limit of the calibration range shall be analyzed without internal standard to evaluate whether relevant amounts of the unlabeled analyte ions appear as isotopically-labeled compound fragments that could impact quantitation.

7.4.3. Evaluating Interferences from Other Commonly Encountered Analytes

For all methods other than immunoassays, it is necessary to evaluate other analytes which may be expected to be present in case samples for their potential to interfere with the method's analytes. For example, a method developed to analyze blood for cocaine must evaluate whether other common drugs of abuse, metabolites, and structurally-similar compounds interfere with the assay. Likewise, a headspace GC-FID method developed for ethanol must evaluate whether other common volatile organic compounds interfere with the assay.

This evaluation is accomplished by analyzing fortified matrix samples, previously analyzed case samples, or neat reference materials of the potential interference(s) at high therapeutic or lethal concentrations, depending on the analyte, the matrix, and the laboratory's mission. The most common drugs/metabolites encountered in the laboratory must be included in the evaluation together with other common drugs within the classification, where appropriate.

7.5. Ionization Suppression/Enhancement

The enhancement or suppression of analyte[4] ionization resulting from the presence of co-eluting compounds is a phenomenon commonly encountered in liquid chromatography/mass spectrometry (LC-MS) applications.

When average suppression or enhancement exceeds ±25% or the % CV of the suppression or enhancement exceeds 15%, a laboratory must demonstrate that there is no impact on other critical validation parameters. For example, suppression or enhancement of ionization is most likely to impact the limit of detection of a qualitative method. Likewise, the limit of detection, the limit of quantitation, and bias may be affected by ionization suppression or enhancement in quantitative methods. The influence on the above parameters shall be assessed by increasing the number of different sources of blank matrices used in their evaluation.

Ionization suppression/enhancement shall be evaluated using either of the approaches that follow.

7.5.1. Post-column Infusion to Assess Ionization Suppression/ Enhancement

This approach provides information on retention times where ionization suppression/enhancement occurs. It is useful for method development, as well as to assess the amount of ionization suppression or enhancement for LC-MS based confirmation methods. Solutions at both low and high concentrations of the analyte are individually infused with a syringe pump into the eluent from the column via a post-column tee-connection and a constant baseline signal for the analyte of interest is monitored. Whenever possible, a minimum of 10 different processed blank matrix samples that are representative of the quality of samples typically encountered in casework are injected into the LC-MS during infusion of the solutions.[5] If there is any considerable suppression or enhancement (>25%) of the infused analyte signal at the

[4] Laboratories must also assess the impact of ionization suppression or enhancement on the method's internal standards unless their daily QA/QC practices address changes in internal standard response.

[5] Additional matrix samples may be required in postmortem toxicology given the variety of sample conditions typically encountered in this work.

retention time of the analyte, then modification of the chromatographic system or the sample preparation may be required to minimize the effect of ionization suppression or enhancement.

7.5.2. Post-Extraction Addition Approach to Assess Ionization Suppression/Enhancement

This approach yields a quantitative estimation of ionization suppression/enhancement. It is useful for assessing the amount of ionization suppression or enhancement for LC-MS based quantitative methods. Two different sets of samples are prepared, and the analyte peak areas of neat standards are compared to matrix samples fortified with neat standards after extraction or processing.

Set one consists of neat standards prepared at two concentrations – one low and one high. Each of these neat standards is injected a minimum of six times to establish a mean peak area for each concentration.

Set two consists of a minimum of 10 different matrix sources, whenever possible.[6] Each matrix source is extracted in duplicate. After the extraction is complete, each matrix sample is then reconstituted/fortified with either the low or high concentration neat standard.

The average area of each set ($\overline{X}$) is used to estimate the suppression/ enhancement effect at each concentration as follows:

$$Ionization\ suppression\ or\ enhancement\ (\%) = \left(\frac{\overline{X}\ Area\ of\ Set\ 2}{\overline{X}\ Area\ of\ Set\ 1} - 1 \right) x100$$

Two ionization suppression or enhancement percentages will be established – one at the low concentration and one at the high concentration.

[6] Additional matrix samples may be required in postmortem toxicology given the variety of sample conditions typically encountered in this work.

7.6. Limit of Detection

Limit of detection (LOD) studies shall be carried out for all methods. There are a number of different approaches for determining the LOD. Select the approach that provides the most reasonable estimation of the detection limit given the analytical instrumentation (or lack thereof) utilized in the method.

A method's LOD incorporates instrumental performance, as well as the sample matrix and inherent procedural limitations. Therefore, the LOD must be assessed over multiple runs using fortified matrix samples from at least three different sources of blank matrix, unless otherwise indicated below. Further, when possible, it is necessary to ensure the defined LOD still satisfies the necessary parameters for identification. For example, matching of a mass spectrum to a reference spectrum within an acceptable match factor can only be achieved by experimental determination of LOD rather than theoretical calculation.

The LOD must be determined by one of the following approaches.

7.6.1. Estimating LOD for a Non-Instrumental Method

This approach is most often used when screening for the presence or absence of a specified analyte or class of analytes (e.g., color tests). To estimate the LOD for a visual, non-instrumental method, samples fortified with decreasing concentrations of analyte are analyzed over a minimum of three runs. When possible, multiple analysts should be involved in estimating the LOD using this approach. The lowest concentration of analyte that yields a positive result on all runs is considered the LOD.

7.6.2. Using the Lowest Non-Zero Calibrator as the LOD

This technique is useful for quantitative methods. In some instances, it may be sufficient to define the LOD as the value of the lowest non-zero calibrator. A minimum of three samples per run of the lowest calibrator shall each be analyzed over three runs to demonstrate that all detection and identification criteria are met. If desired, it is acceptable to use the same calibrator replicates used to establish the calibration model (Section 7.2) for some of the samples used for this method, but additional samples/replicates will be needed to meet the minimum of nine data points.

7.6.3. Using the Decision Point Concentration as the LOD

This technique is useful for qualitative and quantitative methods. In some instances, it may be sufficient to define the LOD as the value of an administratively-defined decision point. For example, a laboratory may choose to define a method's LOD for ethanol as 0.02 g/dL for blood based on the laboratory's administratively defined decision point for reporting this analyte, even though a lower LOD is analytically achievable. Likewise for an immunoassay, a laboratory may choose to use the decision point concentration (that has demonstrated appropriate precision (Section 7.1.2.1)) as the assay's LOD. A minimum of three samples per run of a fortified matrix sample at the concentration of the decision point shall be analyzed over three runs to demonstrate that all detection and identification criteria are met. The data generated for the precision at the decision point concentration in immunoassays will suffice for this LOD requirement.

7.6.4. Estimating LOD Using Background Noise

These approaches for determining LOD are only useful for instrumental methods that demonstrate background noise. A minimum of three different blank source matrices must be used. For example, if the assay is to be used for postmortem blood samples, three independent representative postmortem blood sources are needed.

7.6.4.1. Estimating LOD Using Reference Materials

Three (or more) sources of blank matrix samples fortified at decreasing concentrations are analyzed in duplicate (two separate samples) for at least three runs. The LOD is considered the lowest concentration that 1) yields a reproducible instrument response greater than or equal to three times the noise level of the background signal from the negative samples and 2) achieves acceptable predefined detection and identification criteria (e.g., retention time, peak shape, mass spectral ion ratios).

While it may be possible to visually assess the signal to noise ratio, such an approach is subjective. Therefore, calculate the signal-to-noise ratio or use instrumentation software to determine the ratio. If manually calculated, the

signal is defined as the height response of the analyte peak and the noise is defined as the amplitude between the highest and lowest point of the baseline in an area around the analyte peak. Each replicate shall be independently evaluated.

$$Signal\text{-}to\text{-}Noise = \frac{height\ of\ analyte}{amplitude\ of\ noise}$$

7.6.4.2. Estimating LOD Using Statistical Analysis of Background

To determine the LOD using this approach, a minimum of three sources of blank matrix samples are analyzed in duplicate (two separate samples) over at least three runs. The average and standard deviation (*s*) of the signal (e.g., integrated area of signal at the analyte's retention time) from all negative samples is calculated. Likewise, fortified samples of decreasing concentration are analyzed in duplicate over the course of at least three runs. The lowest concentration of a fortified sample that consistently yields a signal greater than the average signal of the negative samples ($\overline{X}$) plus 3.3 times the standard deviation is identified as the LOD:

$$LOD = \overline{X} + 3.3s$$

7.6.5. Estimating LOD Using a Linear Calibration Curve

This technique is useful for any quantitative method that follows a linear calibration model. A minimum of three calibration curves are constructed across the working range of the analytical method over different runs. The LOD can be estimated from the standard deviation of the *y* intercept (s_y) and the average slope (Avg_m) as:

$$LOD = (3.3\ s_y)/Avg_m$$

7.7. Limit of Quantitation

Limit of quantitation (LOQ) studies shall be carried out for all quantitative methods. There are a number of different approaches for determining a method's LOQ. Select the approach that provides the most reasonable estimation of the quantitation limit given the analytical instrumentation

utilized in the method. A method's LOQ incorporates instrumental performance, as well as the sample matrix and inherent procedural limitations. The LOQ must be assessed over multiple runs using fortified, blank matrix samples from at least three different sources of blank matrix, unless otherwise indicated below.

7.7.1. Using the Lowest Non-Zero Calibrator as the LOQ

In some instances, it may be sufficient to define the LOQ as the value of the lowest non-zero calibrator. A minimum of three samples per run of the lowest calibrator shall be analyzed over three runs to demonstrate that all detection, identification, bias, and precision criteria are met. If desired, it is acceptable to use the same calibrator replicates used to establish the calibration model (Section 7.2) for some of the samples used for this method, but additional samples/replicates will be needed to meet the minimum of nine data points.

7.7.2. Using Decision Point Concentration as the LOQ

In some instances, it may be sufficient to define the LOQ as the value of an administratively-defined decision point. For example, a laboratory may choose to define a method's LOQ for GHB as 5 mg/L for antemortem blood based on the laboratory's administratively-defined decision point for reporting this analyte, even though a lower LOQ is analytically achievable. The concentrations used for this approach must remain within the previously established calibration curve. A minimum of three samples per run of a fortified matrix sample at the concentration of the decision point shall be analyzed over three runs to demonstrate that all detection, identification, bias, and precision criteria are met.

7.7.3. Estimating LOQ Using Reference Materials

Three (or more) sources of blank matrix samples fortified at decreasing concentrations are analyzed in duplicate (two separate samples) over a minimum of three runs. The concentrations used for this approach must remain within the previously established calibration curve. The lowest concentration that is capable of achieving acceptable detection, identification, bias, and precision criteria in all three fortified samples is considered the estimated LOQ.

8. Additional Validation Parameters

In certain instances, it is important to evaluate additional validation parameters, if applicable. These include analyte stability when the matrix is frozen and thawed, processed sample stability, and the effect of sample dilution on bias and precision. A laboratory shall include these parameters in its validation plan and determine if they are applicable to the analytical method or if they are already addressed through other means (i.e., quality assurance practices, published references). The laboratory validation plan must include documentation of this evaluation.

8.1. Dilution Integrity

The effect of sample dilution must be determined during validation of quantitative methods if this is a routine practice within the laboratory. At times, this may be due to low specimen volume requiring the sample or assay to be adjusted appropriately. In other instances, excessively high concentrations that are above the established calibration range may be encountered. To bring the analyte concentration within the validated concentration range, the laboratory procedure may allow for reanalysis after dilution of the sample.

If dilution of a sample is allowed because of high analyte concentration or low sample volume, then the laboratory must evaluate the effect of dilution on the method's bias and precision. This is accomplished by repeating bias and precision studies at common dilution ratios (e.g., 1:2, 1:10, 1:50) utilized by the laboratory and determining if performance criteria are still met.

8.2. Stability

Analyte stability may be affected by a number of variables, including storage conditions and sample processing. Stability experiments shall be designed and carried out to address situations normally encountered in laboratory operations, unless analyte stability is already addressed through other means (i.e., quality assurance practices, published references). All stability determinations shall include a set of samples prepared from reference materials. The reference materials are used to prepare fortified samples of the analyte(s) at both low and high concentrations in each matrix that will be analyzed in the method. It is important that a large enough volume of each of these fortified samples is prepared in order to complete the studies used in the sections below. These fortified samples shall initially be analyzed in triplicate to establish

time zero responses. The average time zero response for each set of samples is compared to the average signals from each of the following stability studies. Linear regression of the average signal (e.g., peak areas[7] *or* ratios of peak area of analyte to internal standard) versus time will allow for an assessment of trends.

8.2.1. Stability – Freeze/Thaw[8]

If it is part of a laboratory's standard practice to freeze samples prior to analysis and there are no published data to rely upon, analyte stability shall be determined after three freeze and thaw cycles. The above fortified samples (Section 8.2) are aliquoted into a minimum of three separate storage containers per concentration and then frozen at the intended storage temperature for 24 hours. This is followed by an unassisted thaw at room temperature. When completely thawed, the first set of samples shall be analyzed in triplicate, while the others are refrozen for 12 to 24 hours under the same conditions. The freeze/thaw cycle and analysis shall be repeated two more times. The analyte will be considered as stable until the average signal (e.g., peak area *or* ratios of peak area of analyte to internal standard) compared to the time zero average signal falls outside of the method's acceptable bias. For example, if the method bias is ±10% and the time zero average signal is 100,000, the analyte is considered stable until the average signal falls outside of the 90,000 – 110,000 range.

8.2.2. Stability – Processed Sample

Circumstances may arise in which samples that have undergone routine preparation for instrumental analysis cannot be immediately analyzed. It may be necessary to run the sample the following day or later. In these instances, it is important to evaluate the length of time a processed sample can be maintained before it undergoes unacceptable changes, preventing reliable analyte detection, identification, or quantitation.

[7] When monitoring peak areas, the instrument's response must be constant over several days for reliable interpretation of the data.

[8] It is recognized that freeze/thaw and storage stability studies in solid samples (e.g., hair, tissues, food products) may not be possible by fortification due to the nature of these matrices. Caution should be employed in interpreting results of solid samples when stability information is not available.

Typically processed samples fortified per Section 8.2 are combined per concentration and then divided into different autosampler vials. As indicated above, the first vials of each concentration are immediately analyzed in triplicate to establish the time zero responses. All remaining vials are stored in a manner that they would typically be stored during routine analysis (e.g., refrigerated, at room temperature on autosampler). The remaining vials are then analyzed in triplicate at different time intervals. Average responses at each time interval are compared to the time zero responses. The analyte will be considered stable until the average signal (e.g., peak area *or* ratios of peak area of analyte to internal standard) compared to the time zero average signal falls outside of the method's acceptable bias. For example, a method's bias limit is ±15% and the time zero average signal is 100,000. Processed samples in different autosampler vials are analyzed repeatedly up to 72 hours. The processed sample's analyte is considered stable until the average signal falls outside of the 85,000 – 115,000 range.

9. Required Revalidation of Previously Validated Methods

Modifications to a validated method require evaluation to confirm that the changes do not have an adverse effect on the method's performance. The decision regarding which performance characteristics require additional validation is based on logical consideration of the specific parameters likely to be affected by the change(s). These changes may include, but are not limited to:

- Analytical conditions
- Instrumentation
- Sample processing
- Data software

For example, changes of extraction solvent or buffer may affect linearity, interferences, LOQ, precision, and bias. A change of the analytical column stationary phase or a change in mobile phase composition may affect linearity and interferences. Further, consideration should be given to conducting parallel studies with known or proficiency samples utilizing both a previously validated method and the modified method in order to evaluate the effects of the changes. The goal is to demonstrate the impact the changes have on the performance of the previously validated procedure.

Laboratories using methods that were validated prior to the promulgation of these minimum standards must demonstrate and document that these methods are fit for use under these standards. These methods will likely have sufficient historical calibration and control data, as well as previously analyzed casework sample results, that can be used to address a number of the required validation parameters. When sufficient data are absent to fulfill these minimum standards, appropriate studies must be conducted to ensure compliance with this document.

10. Documentation Requirements for Method Validation

Record keeping is an essential part of laboratory operating procedures and is a key component of method validation. The data generated during method validation studies must be maintained and available for audits, reviews, or inspections. These records must be organized for easy retrieval and review.

Method validation records must include a summary of the validation studies conducted and their results. The format of this summary report may be a brief bulleted report or table summary format to facilitate a swift review of validation studies. The summary shall minimally include the following:

- Scope
- Validation plan
- Description of all the parameters evaluated. If any of the parameters were not evaluated, then the reason must be stated or justified.
- Sample preparation steps to include concentrations and matrices
- Raw data or reference to where the raw data are stored
- Results and calculations
- Conclusions
- References
- Documentation of management review and approval

It is important that the validation records contain specific details regarding the studies conducted, including:

- Individuals involved in the method validation
- Specific instrumentation
- Dates

Method validation documentation must also include a copy of the newly developed analytical method or a reference to its location. Further, it is

recommended that validation documentation be retained for a minimum of 10 years after the method is retired.

11. Efficiency with Validation

It is recognized that method validation is a time-consuming, expensive, but essential endeavor. Keep in mind that some validation experiments may be conducted concurrently with the same fortified samples. Appendices C, D, and E present example approaches to assist in streamlining validation experiments.

12. References

Araujo P, Key Aspects of Analytical Method Validation and Linearity Evaluation, *J Chromatogr B*, 877, (2009), 2224-2234.

Bressolle F, Bromet-Petit M, Audran, M, Validation of Liquid Chromatographic and Gas Chromatographic Methods, Applications to Pharmacokinetics, *J Chromatogr B*, 686, (1996), 3-10.

Bruce P, Minkkinen, Riekkola M, Practical Method Validation: Validation Sufficient for an Analytical Method, *Mikrochim Acta*, 128 (1998), 93-106.

Clinical and Laboratory Standards Institute, C50-A Mass Spectrometry in the Clinical Laboratory: General Principles and Guidance; Approved Guideline, Vol 27, No 24 (2007).

Corley J, Best Practices in Establishing Detection and Quantification Limits for Pesticides Residues in *Foods in Handbook of Residue Analytical Methods for Agrochemicals*, John Wiley & Sons Ltd, Chichester (2003).

Drummer O, Requirements for Bioanalytical Procedures in Postmortem Toxicology, *Anal Bioanal Chem*, 388 (2007), 1495-1503.

Eurachem Guide, The Fitness for Purposes of Analytical Methods, (1998).

Health Sciences Authority, Guidance Notes on Analytical Method Validation: Methodology, (Sept. 6, 2004).

Hubert Ph, Nguyen-Huu J-J, Boulanger B et al., Harmonization of Strategies for the Validation of Quantitative Analytical Procedures ASFSTP Proposal – Part I, *J Pharm Biomed Anal*, 36 (2004), 579-586.

Hubert Ph, Nguyen-Huu J-J, Boulanger B et al., Harmonization of Strategies for the Validation of Quantitative Analytical Procedures ASFSTP Proposal – Part II, *J Pharm Biomed Anal*, 45 (2007), 70-81.

Hubert Ph, Nguyen-Huu J-J, Boulanger B et al., Harmonization of Strategies for the Validation of Quantitative Analytical Procedures ASFSTP Proposal – Part III, *J Pharm Biomed Anal*, 45 (2007), 82-96.

International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), ISO/IEC Guide 99:2007, Organization for Standardization (ISO)/International Electrotechnical Commission (IEC), Geneva, 2007.

Kushnir M M, Rockwood A L, Nelson G J, et al., Assessing Analytical Specificity in Quantitative Analysis Using Tandem Mass Spectrometry, *Clin Biochem*, 12:003, (2004), 319-327.

Liang H R, Foltz R L, Meng M, Bennett P, Ionization enhancement in atmospheric pressure chemical ionization and suppression in electrospray ionization between target drugs and stable-isotope-labeled internal standards in quantitative liquid chromatography/tandem mass spectrometry, *Rapid Comm Mass Spec*, 17 (2003), 2815-2821.

Matuszewski B K, Constanzer M L, Chavez-Eng C M, Strategies for the Assessment of Matrix Effects in Quantitative Bioanalytical Methods Based on HPLC-MS/MS, *Anal Chem*, 75: 13, (2003), 3019-3030.

Niedbala R S and Gonzalez J M, Immunoassays in Clarke's Analysis of Drugs and Poisons, 4th edition, A C Moffat, M D Osselton, B Widdop, and J Watts Eds. Pharmaceutical Press: London (2011).

Peters F, Maurer H H, Bioanalytical Method Validation and Its Implications for Forensic and Clinical Toxicology – A Review, *Accred Qual Assur*, 7, (2002), 441-449.

Peters FT, Method Validation in *Applications of LC-MS in Toxicology*, Pharmaceutical Press, London (2006).

Peters F, Drummer O, Musshoff F, Validation of New Methods, *Forensic Sci Int*, 165 (2007), 216-224.

Remane D, Meyer MR, Wissenbach DK, Mauer HH, Ion suppression and enhancement effects of co-eluting analytes in multi-analyte approaches: systematic investigation using ultra-high performance liquid

chromatography/mass spectrometry with atmospheric pressure chemical ionization or electrospray ionization, *Rapid Comm Mass Spec*, 24 (2010), 3103-3108.

Shah V, Midha K, Dighe S, et al, Analytical Methods Validation: Bioavailability, Bioequivalence and Pharmacokinetics Studies, *Pharm Res*, 9:4, (1992), 588-592.

Shah V, Midha K, Findlay J, et al., Bioanalytical Method Validation – A Revisit with a Decade of Progress, *Pharm Res*, 17:12 (2000), 1551-1557.

Shultz E K, Analytical Goals and Clinical Interpretation of Laboratory Procedures in Tietz Textbook of Clinical Chemistry, 2nd edition, C A Burtis and E R Ashwood, Eds. W.B. Saunders Company: Pennsylvania (1994).

Stockl D, D'Hondt H, Thienpont L M, Method Validation Across the Disciplines – Critical Investigation of Major Validation Criteria and Associated Experimental Protocols, *J Chromatogr B*, 877, (2009), 2180-2190.

Thompson M, Ellison S L R, Wood R. Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis, IUPAC Technical Report, *Pure Appl Chem*, 74:5, (2002), 835-855.

U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for Industry, Bioanalytical Method Validation, (2001).

Van Eeckhart A, Lanckmas K, Sarre S, et al., Validation of Bioanalytical LC-MS/MS Assays: Evaluation of Matrix Effects, *J Chromatogr B*, 877, (2009), 2198-2207.

Viswanathan C T, Bansal S, Booth B, et al., Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays, *Pharm Res*, 24:10, (2007), 1962-1973.

Viswanathan C T, 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*, 9:1, (2007), E30-E42.

Appendix A: Quantitation of Drug X in Blood Validation Example

The following is an example of some of the validation steps outlined in this document. It is not intended to provide specific guidance for any particular method.

In this example, assume a laboratory validated a LC/MS/MS method for a new opiate, Drug X, in whole blood.

Create Validation Plan (Section 5)

Before starting the validation experiments, the laboratory prepared the validation plan. In the plan, they specified that an existing SPE procedure, already used for the extraction of other opiates, would be relied upon for extracting Drug X (Section 4). Further, instrument conditions were previously optimized (Section 4), so those conditions were also listed in the plan (not shown). As this is a quantitative procedure, the validation parameters listed in Table A1 were assessed against the laboratory's pre-defined acceptance criteria.

Table A1: Validation parameters to be assessed

Parameter:	Acceptance Criteria:
Bias	Must not exceed ± 20%
Calibration Model	10 – 1000 ng/mL (linear model desired)
Carryover	Carryover after highest calibrator does not exceed 10% of signal of lowest calibrator
Interference Studies	No interfering signal from matrix, internal standard, common drugs of abuse (including other common opiates/metabolites), OTC drugs, and prescription medications
Ionization Suppression/ Enhancement	<25% suppression or enhancement and <15% CV due to matrix (if not, evaluate impact on LOD, LOQ, and Bias)
Limit of Detection	Must be 10 ng/mL or lower
Limit of Quantitation	Must be 10 ng/mL or lower
Precision	% CV must not exceed 20%
Dilution Integrity	Bias and precision criteria must be met with dilution of samples. Dilution ratios evaluated will depend on linear range of final calibration curve.
Processed Sample Stability	Evaluate length of time that analyte in extracted samples stored at room temperature on autosampler remains stable

Interference Studies (Section 7.4)

Ten independent sources of blank whole blood were secured from previously analyzed cases to evaluate matrix interferences (Section 7.4.1). The blank matrix samples were extracted *without* the addition of internal standard (d3-Drug X) and

analyzed using the newly developed method. No interferences at the retention time for Drug X were noted after analysis of the blank whole blood samples.

The laboratory randomly selected one of the blank matrix samples, added d3-Drug X to the sample (250 ng/mL), extracted the sample, and analyzed it. This was to demonstrate that the internal standard would not interfere with the signal for Drug X (Section 7.4.2). Likewise, another random blank matrix sample was fortified with Drug X at 2000 ng/mL and analyzed *without* internal standard. This was to evaluate whether the unlabeled analyte ions interfere with the signal for d3-Drug X. The results demonstrated no interferences between the analyte and internal standard.

Lastly, to evaluate interferences from other commonly encountered analytes (Section 7.4.3), the laboratory injected neat solutions diluted in mobile phase to a concentration of 5000 ng/mL (or higher) of all common opiates and metabolites observed in their casework, other common recreational drugs of abuse and their metabolites, other common prescription medications and their metabolites, and common over-the-counter medications and their metabolites. Table A2 shows how the laboratory efficiently prepared these solutions into four injection standards. The laboratory observed no interference for the signal of Drug X or d3-Drug X from any of these compounds.

Table A2: Example drugs/metabolites used in interference study

Injection Mix	Included Drugs/Metabolites (5000 ng/mL unless noted otherwise)
Opiates and Related	codeine, morphine, heroin, 6-acetylmorphine, hydrocodone, hydromorphone, oxycodone, oxymorphone, levorphanol, meperidine, methadone, tramadol, fentanyl
Drugs of Abuse	amphetamine, cocaine, benzoylecgonine, ecgonine methyl ester, methamphetamine, PCP, MDA, MDMA, THC, THC-COOH
Prescription Drugs	antidepressants (amitriptyline, imipramine, doxepin, amoxapine, trazodone, bupropion, fluoxetine, sertraline, citalopram), benzodiazepines (alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam), antiarrhythmics (verapamil, diltiazem, lidocaine), barbiturates at 500,000 ng/mL (amobarbital, butalbital, pentobarbital, phenobarbital), other CNS depressants (zopiclone, buspirone, zolpidem)
OTC Drugs	antihistamines (diphenhydramine, doxylamine, chlorpheniramine), analgesics at 500,000 ng/mL (acetaminophen, ibuprofen), antitussive (dextromethorphan)

Calibration Model (Section 7.2) and Carryover (Section 7.3)

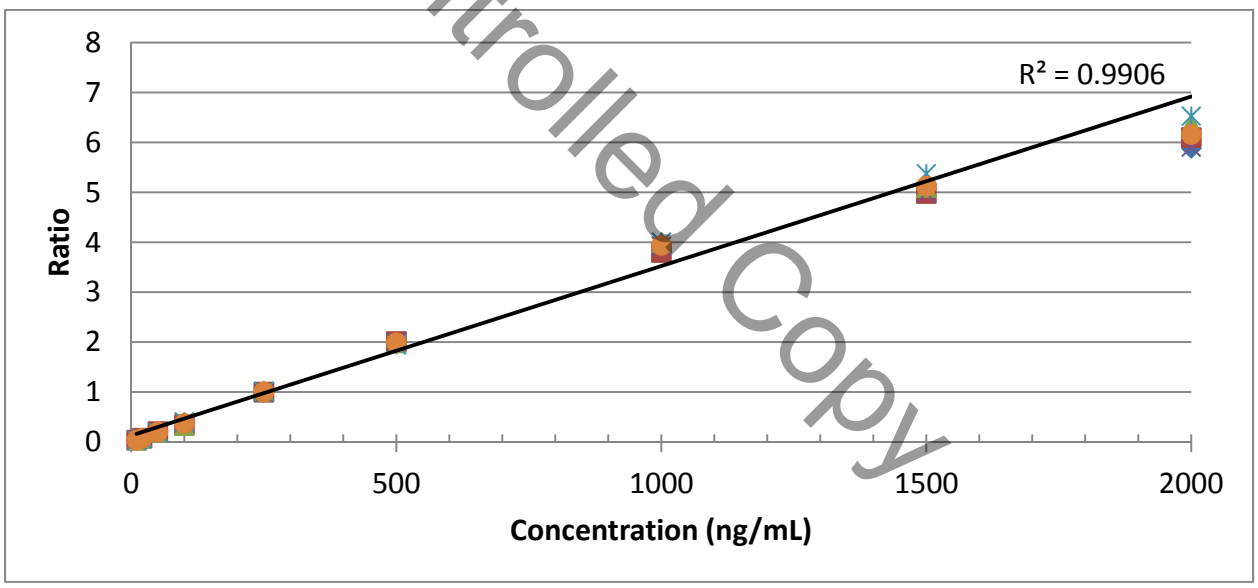
The laboratory indicated a desire for the method's calibration model to be linear and include the range of 10 – 1000 ng/mL. However, to evaluate if the method could exceed this range, the calibration samples were prepared in blank blood at the concentrations of 10, 20, 50, 100, 250, 500, 1000, 1500, and 2000 ng/mL. Each calibrator was analyzed once per run in five separate runs (Table A3). An extracted matrix blank was analyzed after each calibrator to evaluate carryover at each concentration. The data of all runs were combined into a single calibration curve. It was noted that carryover was not present for Drug X or the internal standard in any of the extracted blank matrix samples that followed the calibrators in the range of 10 – 1500 ng/mL; however, a small amount of carryover for Drug X was observed in two of the five blank matrix samples that followed the 2000 ng/mL calibrator. The integrated areas of Drug X in these two samples were less than 10% of the smallest area of the lowest (10 ng/mL) calibrator, so the carryover from the 2000 ng/mL calibrator was deemed acceptable.

The first evaluation of these data suggested that linearity may break off above 1000 ng/mL (Table A3 and Figure A1). A residual plot was used to further evaluate these data (Figure A2).

Table A3: Calibration curve data

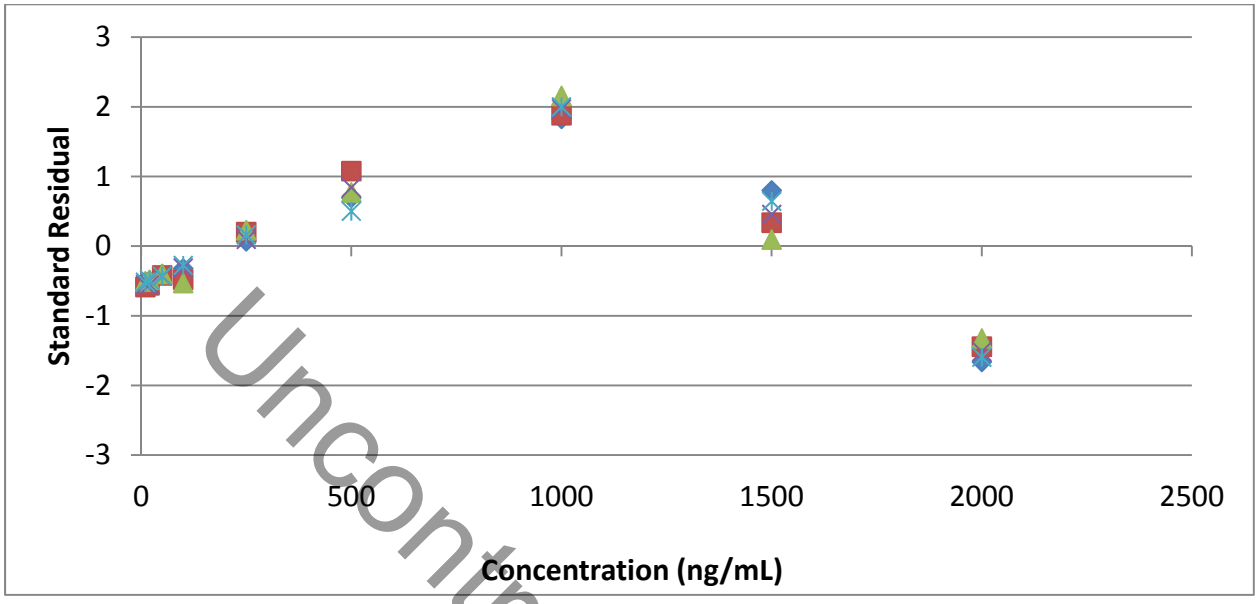
Conc (ng/mL)	Run 1			Run 2			Run 3			Run 4			Run 5		
	Peak Area		Ratio	Peak Area		Ratio	Peak Area		Ratio	Peak Area		Ratio	Peak Area		Ratio
	Drug X	Int Std		Drug X	Int Std		Drug X	Int Std		Drug X	Int Std		Drug X	Int Std	
10	3951	101310	0.039	4112	100281	0.041	3971	101833	0.039	4319	102831	0.042	3872	101885	0.038
20	7937	105831	0.075	7930	100382	0.079	8254	105821	0.078	7511	104321	0.072	8154	103215	0.079
50	20470	100838	0.203	21753	103588	0.21	20302	102018	0.199	21590	105832	0.204	20266	103927	0.195
100	39703	102328	0.388	38273	108729	0.352	34141	103771	0.329	40784	102731	0.397	41763	103887	0.402
250	100584	100887	0.997	108547	108439	1.001	105062	102800	1.022	100227	100832	0.994	107173	105382	1.017
500	211080	105382	2.003	215146	106772	2.015	207431	104394	1.987	208572	103819	2.009	203887	104665	1.948
1000	412751	103889	3.973	379832	99982	3.799	417110	104382	3.996	395587	100838	3.923	403819	100728	4.009
1500	517537	99872	5.182	477045	95773	4.981	483159	94737	5.1	475616	95333	4.989	525150	97666	5.377
2000	560931	95283	5.887	561994	92100	6.102	593173	93076	6.373	531467	89942	5.909	606638	92886	6.531

Figure A1: Combined calibration curve demonstrating loss of linearity above 1000 ng/mL



The residual plot showed an inverted U-shaped distribution suggesting a non-linear model would be the best calibration model for these data (Figure A2).

Figure A2: Standard residual plot of calibration curve data with an inverted U-shaped distribution



Because the laboratory's preference was to use a linear calibration model, they re-evaluated these data after dropping the 1500 and 2000 ng/mL calibrators. Doing so allowed for their original validation plan requirements (10 – 1000 ng/mL) to still be met. The revised calibration curve appeared to provide a better fit of these data using an unweighted linear model (Figure A3). This was confirmed by the residual plot that showed a random distribution around the zero line suggesting a linear model was the most appropriate for these data (Figure A4).

Figure A3: Revised calibration curve

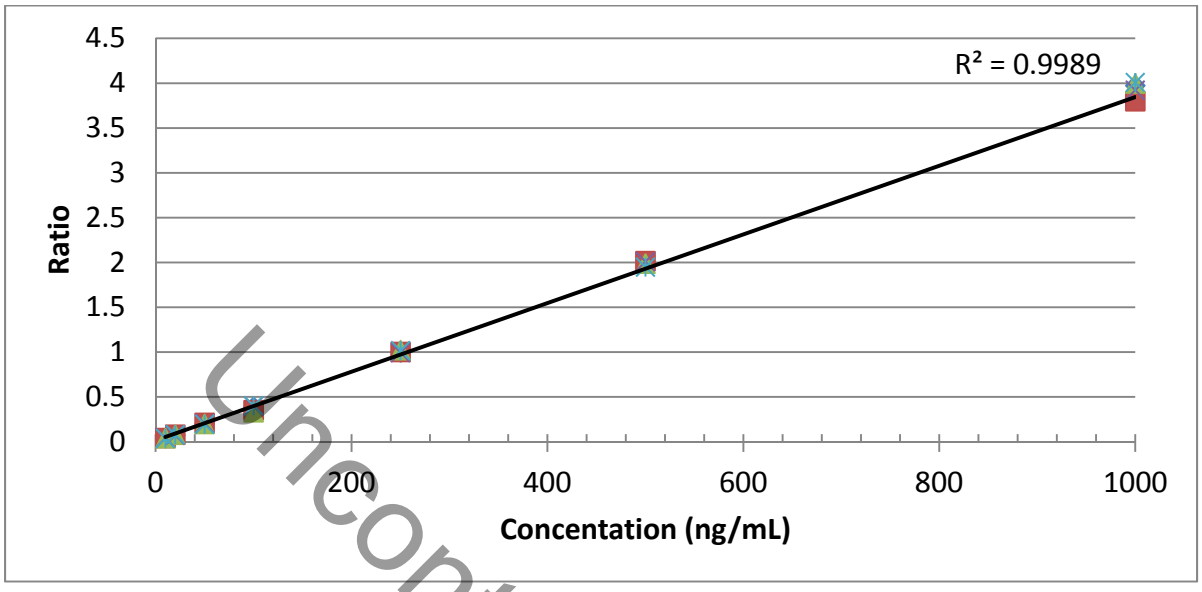
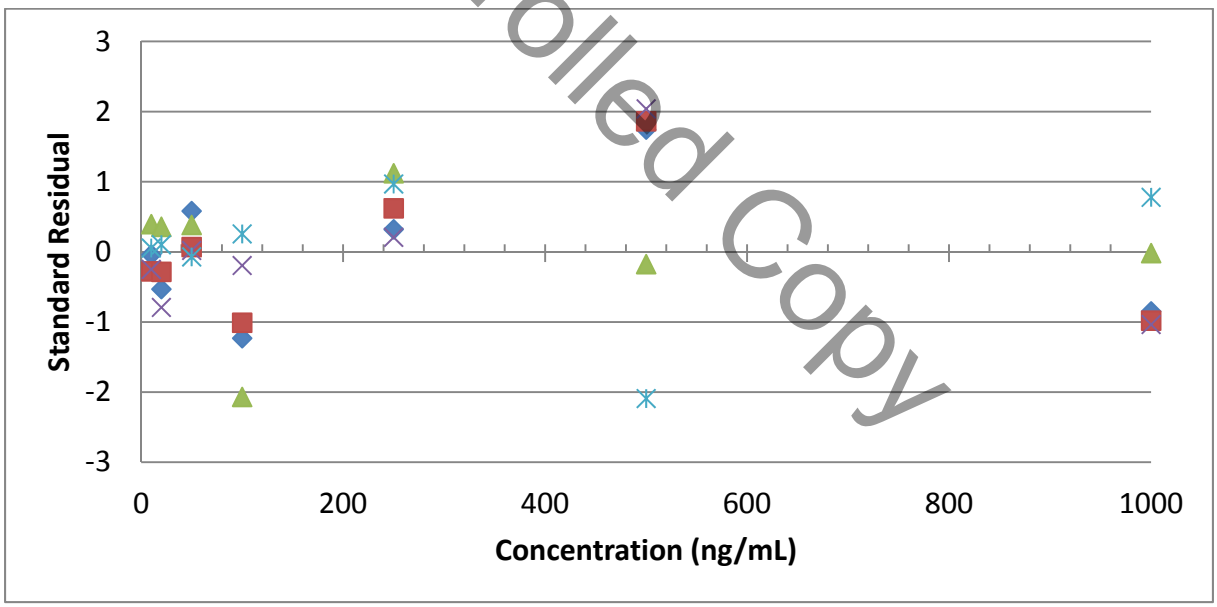


Figure A4: Residual plot of calibration curved data with a random distribution



For future validation experiments, the laboratory used calibrators prepared at 10, 50, 100, 250, 500, and 1000 ng/mL.

Since accurate quantitative results cannot be assumed above 1000 ng/mL, the laboratory knew they would have to re-extract (with dilution) any samples that

exceed 1000 ng/mL. Therefore, they planned to evaluate dilution integrity in ratios up to 1:5 when conducting the bias and precision experiments.

Recall that no carryover was observed up to 1500 ng/mL in the laboratory's study. However, since the working calibration range will end at 1000 ng/mL, the laboratory recognized that accurate quantitative results cannot be achieved above the working range. Therefore, carryover will need to be evaluated in samples that follow those that exceed 1000 ng/mL. If the amount of Drug X in samples following those with concentrations greater than 1000 ng/mL is above the method's LOD, the samples with potential carryover will be re-extracted and analyzed.

Limit of Detection (Section 7.6)

To estimate the LOD, the laboratory chose to utilize the results from their previously generated calibration curve data (Section 7.6.5). Both the slope and *y*-intercept of the individual calibration curves (10 – 1000 ng/mL) were determined in order to calculate the average slope and standard deviation of the *y*-intercept s (Table A4).

Table A4: Slope and y-intercept data from calibration curves

	Slope	*y*-Intercept
Run 1	0.003980	-0.00050
Run 2	0.003828	-0.01543
Run 3	0.004009	-0.01247
Run 4	0.003934	0.00695
Run 5	0.003995	-0.00318
Average	**0.003949**	0.00125
Std Dev	0.000073	**0.01054**

The LOD was calculated using the formula:

3.3 x 0.01054 / 0.003949 = 8.8 ng/mL

Limit of Quantitation (Section 7.7)

The laboratory chose to analyze reference materials to establish their LOQ (Section 7.7.3). Three sources of whole blood were each fortified at 20, 15, and 10 ng/mL. They were extracted and analyzed in duplicate against a freshly prepared calibration curve on three different days. The lowest concentration that was capable of reproducibly providing symmetrical peaks and the minimum mass spectral identification ratios, while maintaining a bias of ±20% and a % CV of <20%, was the 10 ng/mL sample. This concentration was deemed as the method's LOQ and reaffirmed acceptable results at the lowest calibration point.

Bias and Precision (Section 7.1)

To establish the method's bias and precision, the laboratory prepared three pools of fortified matrix samples at the following concentrations: low (30 ng/mL); medium (400 ng/mL); and high (800 ng/mL). Each concentration pool of fortified samples was analyzed in triplicate on five separate days along with a freshly prepared calibration curve (Table A5).

The laboratory calculated the bias (Section 7.1.1) by first determining the mean for each concentration. This resulted in the values listed in Table A6.

From these values, the bias was calculated at each concentration. For example, for the low concentration sample, the bias was determined as:

$$Bias_{Low} = ((28 - 30 / 30) \times 100) = (- 6.7\%)$$

Likewise, the bias for the medium and high concentrations was calculated as 9.3% and -2.4%, respectively.

Table A5: Quantitative results (ng/mL) of bias and precision runs

Low (30 ng/mL)	Run 1	Run 2	Run 3	Run 4	Run 5
Rep 1	32	26	29	26	28
Rep 2	28	24	31	35	25
Rep 3	27	28	27	30	29
Med (400 ng/mL)	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Rep 1	412	435	427	455	444
Rep 2	444	410	419	438	442
Rep 3	422	450	479	452	423
High (800 ng/mL)	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Rep 1	892	793	761	742	820
Rep 2	827	741	729	734	749
Rep 3	850	769	803	720	791

Table A6: Mean concentrations (ng/mL) for bias calculations

Conc (ng/mL)	Calculated Mean	Bias
Low (30)	28	-6.7%
Med (400)	437	9.3%
High (800)	781	-2.4%

Within-run and between-run precisions were calculated using the one-way ANOVA approach (Section 7.1.2.2.3). Using the ANOVA: Single Factor analysis in popular spreadsheet or statistics programs (see Table A7 for Low Concentration), the laboratory was able to obtain values for the mean square within groups for the low concentration and introduced into the appropriate formula as follows:

$$Within - run\ CV(\%) = \left[\frac{\sqrt{MS_{wg}}}{grand\ mean\ for\ each\ concentration}\right] x100$$

$$Within - run\ CV(\%) = \left[\frac{\sqrt{7.933}}{28}\right] x100$$

$$Within - run\ CV(\%) = \left[\frac{2.817}{28}\right] x100$$

$$Within - run\ CV(\%) = 10.1\%$$

Table A7: ANOVA calculations for 30 ng/mL sample

ANOVA: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	87	29	7
Column 2	3	78	26	4
Column 3	3	87	29	4
Column 4	3	91	30.33333	20.33333
Column 5	3	82	27.33333	4.333333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	34	4	8.5	1.071429	0.420175	3.47805
Within Groups	79.33333	10	7.933333			
Total	113.3333	14				

The between-run precision for the low concentration was calculated using the formula and the mean square between groups from the ANOVA table:

$$Between-Run\ CV(\%) = \left[\frac{\sqrt{\frac{MS_{bg} + (n-1) * MS_{wg}}{n}}}{grand\ mean\ for\ each\ concentration}\right] x100$$

$$Between-Run\ CV(\%) = \left[\frac{\sqrt{\frac{8.5 + (3-1) * 7.933}{3}}}{28}\right] x100$$

$$Between-Run\ CV(\%) = \left[\frac{2.845}{28}\right] x100$$

$$Between-Run\ CV(\%) = 10.2\%$$

Using the data for the medium and high concentrations, the ANOVA: Single Factor analysis was conducted on these levels (data not shown) and appropriate values introduced into the formulas to obtain the within-run and between-run precisions. Table A8 lists the calculated results for all concentrations.

Table A8: Precision results

	Low	Medium	High
Within-Run	10.1% CV	4.5% CV	3.9% CV
Between-Run	10.2% CV	4.2% CV	2.2% CV

Ionization Suppression/Enhancement (Section 7.5)

As the instrumental portion of the method involves LC/MS/MS, the laboratory was required to conduct ionization suppression/enhancement experiments. The post-column extraction approach was chosen (Section 7.5.2).

Three sets of samples were prepared for the experiment. Set one consisted of standards prepared in mobile phase at 30 and 800 ng/mL. They were not extracted, but instead simply injected six times each.

Set two was prepared in 10 blank matrix samples. Each blank matrix was from an independent source of blank whole blood from previously analyzed cases.

These were the same ten blank matrix samples used in the interference studies. The blank matrix samples were extracted in duplicate *and then* fortified to 30 and 800 ng/mL with Drug X and 250 ng/mL with d3-Drug X. Each concentration set sample was injected one time each.

Average peak areas for both the Drug X and the d3-Drug X are found in Table A9.

Table A9: Average peak areas from suppression/enhancement experiments

Average Peak Areas			
30 ng/mL		800 ng/mL	
Drug X	d3-Drug X	Drug X	d3-Drug X
13890	110381	330822	112827
11812	102444	303992	105923

Using the above data sets, the laboratory calculated the % ionization suppression/ enhancement for each concentration using the formula:

$$\%\ Ionization\ suppression/enhancement_{Drug\ X\ (Low)} = ((11812 / 13890)-1) \times 100 = (-15.0\%)$$

The negative value suggested some suppression was occurring, but it was less than 25%.

Similarly, the laboratory calculated the % suppression/enhancement for 800 ng/mL and for the internal standard in both sets. The results suggested suppression of -8.1% for Drug X at the 800 ng/mL concentration. Although at the same concentration in both the low and high samples, the d3-Drug X demonstrated ionization suppressions of 7.2% and 6.1%, respectively.

The data were also used to calculate the % CV at each concentration. All % CVs were <14% (data not shown).

Since the average suppression for all analytes did not exceed ±25% and the calculated % CV value was <15%, the variation was considered insignificant. No further work was required for other validation parameters.

Dilution Integrity (Section 8.1)

While the laboratory indicated that a minimum working range for the calibration curve was between 10 and 1000 ng/mL, they anticipated occasional samples that contain Drug X at concentrations above 1000 ng/mL. Their initial attempt to

extend the calibration range to 2000 ng/mL was abandoned when they realized that a non-linear calibration model would be needed. Therefore, they conducted dilution integrity experiments to demonstrate acceptable bias and precision results when samples are diluted in deionized water. They evaluated two dilutions ratios: 1:2 and 1:5.

The laboratory prepared two fortified matrix samples at concentrations of 1600 ng/mL and 3000 ng/mL. The 1600 ng/mL sample was diluted 1:2 before extraction and analysis. Likewise, the 3000 ng/mL sample was diluted 1:5. Both dilution samples were analyzed in triplicate over five different runs, each with a freshly prepared calibration curve. Bias and precision calculations were performed, and results (Table A10) demonstrated comparable values compared to the results obtained without dilution. This provided proof of no detrimental impact when diluting the samples before extraction.

Table A10: Effect of dilution on bias and precision

	1600 ng/mL (1:2 dilution)	3000 ng/mL (1:5 dilution)
Bias	8.2%	9.9%
Within-Run Precision	4.0%	2.9%
Between-Run Precision	4.4%	3.7%

Processed Sample Stability (Section 8. 2)

The laboratory recognized that samples are not always analyzed immediately after extraction due to large batches or unforeseen delays. For example, the instrument may lose communication with its controller, inadvertently shutting down a batch run. Therefore, to evaluate the impact of room temperature storage of processed samples sitting on the autosampler before analysis, the laboratory conducted a stability study on extracted samples. This was achieved by preparing fortified matrix samples at two concentrations, 30 ng/mL and 800 ng/mL. Twelve aliquots of each concentration were extracted. Reconstituted extracts for each concentration were combined and vortexed to ensure adequate mixing. The concentration pool was then divided into 12 autosampler vials and placed on the autosampler. The first vial of each level was injected three times to represent the time zero (t_0) sample. The remaining vials for each concentration

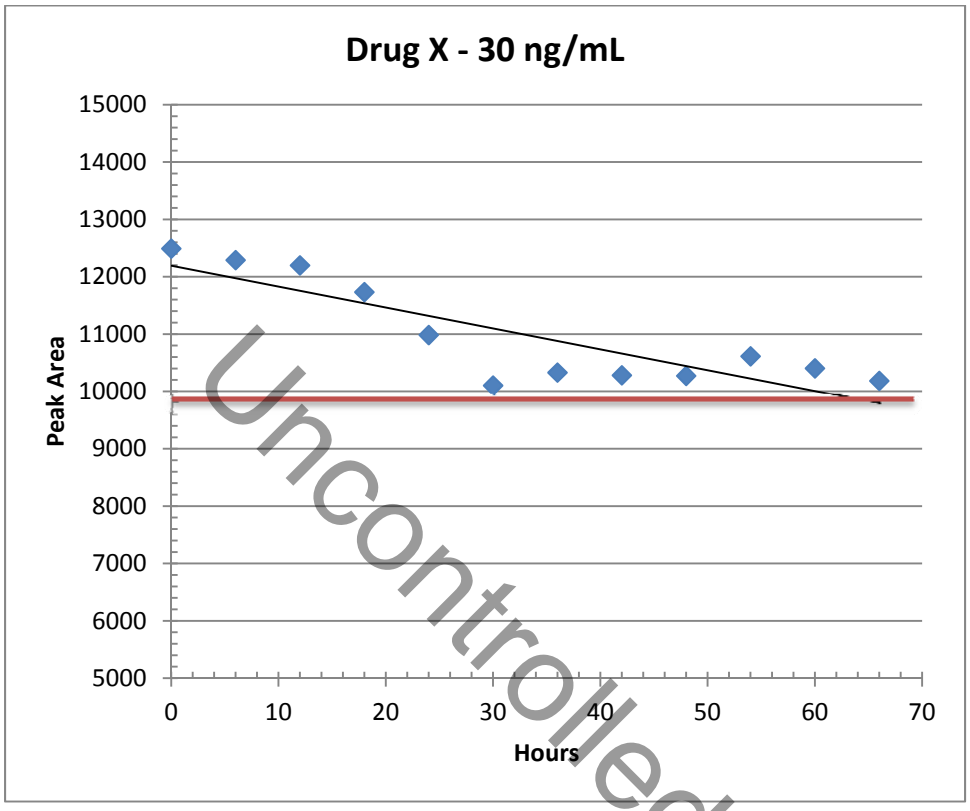
were analyzed in triplicate every six hours up to 66 hours. Analyte signals from the triplicate analyses were averaged and compared to the t_0 signals (Table A11).

Table A11: Average peak areas for processed sample stability study

Time (hr)	Average Peak Area			
	30 ng/mL		800 ng/mL	
	Drug X	d3-Drug X	Drug X	d3-Drug X
0	12490	101832	332554	100423
6	12289	100382	331820	100328
12	12198	100432	330779	101101
18	11732	100733	330246	100987
24	10983	100992	329787	100832
30	10101	101789	326048	100821
36	10328	100904	327238	100234
42	10281	100086	326838	100323
48	10271	100183	315009	99727
54	10612	100309	315772	99421
60	10402	100233	316231	96381
66	10183	100872	315499	94832

By plotting the average peak areas for both Drug X and the internal standard, the laboratory could evaluate the processed samples while they were stored on the autosampler. As their required bias is ±20%, they considered the compounds stable until they saw a decrease (or increase) in signal of more than 20% from the t_0 average peak area. The plot for the 30 ng/mL concentration of Drug X is shown to demonstrate this concept (Figure A5).

Figure A5: Change in Drug X peak area over 66 hours



These data appear to suggest Drug X remained stable within the pre-defined limits for the entire 66-hour period of the study. However, the trend line shows that 66 hours may be the maximum period of time before the samples may need to be re-extracted. It was noted that at the 30-hour mark, stability seemed to have dropped very close to the "instability" point. Since the previously determined bias was actually much better than the ±20% required in their validation plan, the laboratory made a decision to re-extract any samples that remain on the autosampler more than 24 hours.

Documentation of Results (Section 10)

Along with all of the other required documentation listed in Section 10, the laboratory compared the results from the validation studies conducted to the originally defined requirements, as demonstrated in Table A12.

Table A12: Summary of validation results

Parameter:	**Acceptance Criteria:**	**Result:**
Bias	Must not exceed ± 20%	-6.7 to 9.3%
Calibration Model	10 – 1000 ng/mL (linear model desired)	10 – 1000 ng/mL (linear model)
Carryover	Carryover after highest calibrator does not exceed 10% of signal of lowest calibrator	No significant carryover at 2000 ng/mL. Re-extract and analyze samples containing Drug X above the LOD if that sample follows one that exceeds 1000 ng/mL of Drug X.
Interference Studies	No interfering signal from matrix, internal standard, common drugs of abuse (including other common opiates/metabolites), OTC drugs, and prescription medications	No observed interferences from matrix or from common drugs/metabolites
Ionization Suppression/ Enhancement	<25% suppression or enhancement and <15% CV due to matrix (if not, evaluate impact on LOD, LOQ, and Bias)	-8.1 to -15.0%; <14% CV
Limit of Detection	Must be 10 ng/mL or lower	8.8 ng/mL
Limit of Quantitation	Must be 10 ng/mL or lower	10 ng/mL
Precision	% CV must not exceed 20%	Within-run (3.9 to 10.1%) Between-run (2.2 to 10.2%)
Dilution Integrity	Bias and precision criteria must be met with dilution of samples	Using 1:2 and 1:5 aqueous dilutions, bias (8.2 to 9.9%) and precision (within-run (2.9 to 4.0%); between-run (3.7 to 4.4%). Comparable to results obtained without dilution.
Processed Sample Stability	Evaluate length of time that analyte in extracted samples stored at room temperature remains stable	24 hours

Appendix B: Immunoassay Screen of Benzodiazepines in Urine Validation Example

The following is an example of the immunoassay validation steps outlined in this document. It is not intended to provide specific guidance for any particular method.

In this example, assume a laboratory validated an immunoassay kit for its ability to screen urine for benzodiazepines.

Create Validation Plan (Section 5)

Before starting the validation experiments, the laboratory prepared the validation plan. In the plan, it specified that it will use Company ABC's ELISA Immunoassay Kit for Benzodiazepines (Oxazepam) designed with a "cutoff" of 300 ng/mL. The laboratory planned to select its own cutoff concentration (decision point) of 100 ng/mL for the target compound of oxazepam. The sample preparation steps, as well as instrumental settings were listed in the plan. The validation parameters were assessed against the pre-defined requirements listed in Table B1.

Table B1: Validation parameters to be assessed

Parameter:	Desired Limit:
Limit of Detection	Same as decision point (100 ng/mL for oxazepam, lorazepam and alpha-hydroxyalprazolam and 50 ng/mL for alprazolam)
Precision	% CV must not exceed 20%; means ± SD cannot overlap

Precision at the Decision Point (Section 7.1.2.1)

The product brochure listed the cross-reactivities for oxazepam, other benzodiazepines, and their metabolites. An abbreviated list of these cross-reactivities is as follows: oxazepam (100%); nordiazepam (425%); lorazepam (175%); alprazolam (450%); and alpha-hydroxyalprazolam (340%).

Since the assay was to be used to determine the use of the broad class of benzodiazepines, the laboratory was required to verify precision for oxazepam and any other analytes that they chose to screen for using this assay with cross-reactivities less than 100% or with a decision concentration less than that of oxazepam (100 ng/mL). For example, this laboratory decided to use the assay to

screen for lorazepam (decision concentration 100 ng/mL), alprazolam (decision concentration 50 ng/mL), and alpha-hydroxyalprazolam (decision concentration 100 ng/mL). Both oxazepam and alprazolam had to be evaluated for precision at their decision concentration. However, since lorazepam and alpha-hydroxyalprazolam had cross-reactivities greater than 100% and the same decision point as oxazepam (100 ng/mL), precision studies were not required for these assays.

The laboratory prepared three pools of oxazepam-fortified matrix samples at the following concentrations: 50 ng/mL (50% below); 100 ng/mL (decision point); and 150 ng/mL (50% above). Each of the fortified sample sets was analyzed in triplicate on five separate days. The results are shown in Table B2.

Table B2: Results (optical density) of precision runs for the oxazepam sample sets. Each result is the signal obtained from the analysis of the fortified matrix sample.

50 ng/mL	Run 1	Run 2	Run 3	Run 4	Run 5
Rep 1	1.729	1.813	1.708	1.699	1.793
Rep 2	1.771	1.679	1.690	1.723	1.746
Rep 3	1.642	1.694	1.714	1.677	1.681
Mean	**1.714**	**1.729**	**1.704**	**1.700**	**1.740**
Grand Mean			**1.717**		
Std Dev			**0.047**		
100 ng/mL	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Rep 1	1.551	1.566	1.536	1.559	1.568
Rep 2	1.552	1.498	1.514	1.580	1.544
Rep 3	1.529	1.563	1.541	1.508	1.571
Mean	**1.544**	**1.542**	**1.530**	**1.549**	**1.561**
Grand Mean			**1.545**		
Std Dev			**0.024**		
150 ng/mL	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Rep 1	1.089	1.110	0.998	1.001	1.010
Rep 2	1.112	1.057	1.102	0.991	1.092
Rep 3	1.104	1.033	1.048	0.983	1.101
Mean	**1.102**	**1.067**	**1.049**	**0.992**	**1.068**
Grand Mean			**1.055**		
Std Dev			**0.049**		

The result for the 100 ng/mL decision point concentration when considering the standard deviation of the measurement plus or minus two standard deviations

(1.545 ± (2 x 0.024)) was between 1.497 and 1.593. This range did not overlap with the ranges calculated for the 50 ng/mL or 150 ng/mL samples.

The % CV for each concentration was 5.8%, 1.5%, and 4.6%, respectively, well below the requirement to not exceed 20%.

Similar experiments were conducted for alprazolam at the 50 ng/mL decision point, as well as concentrations ±50% of the decision point (data not shown).

Limit of Detection (Section 7.6.3)

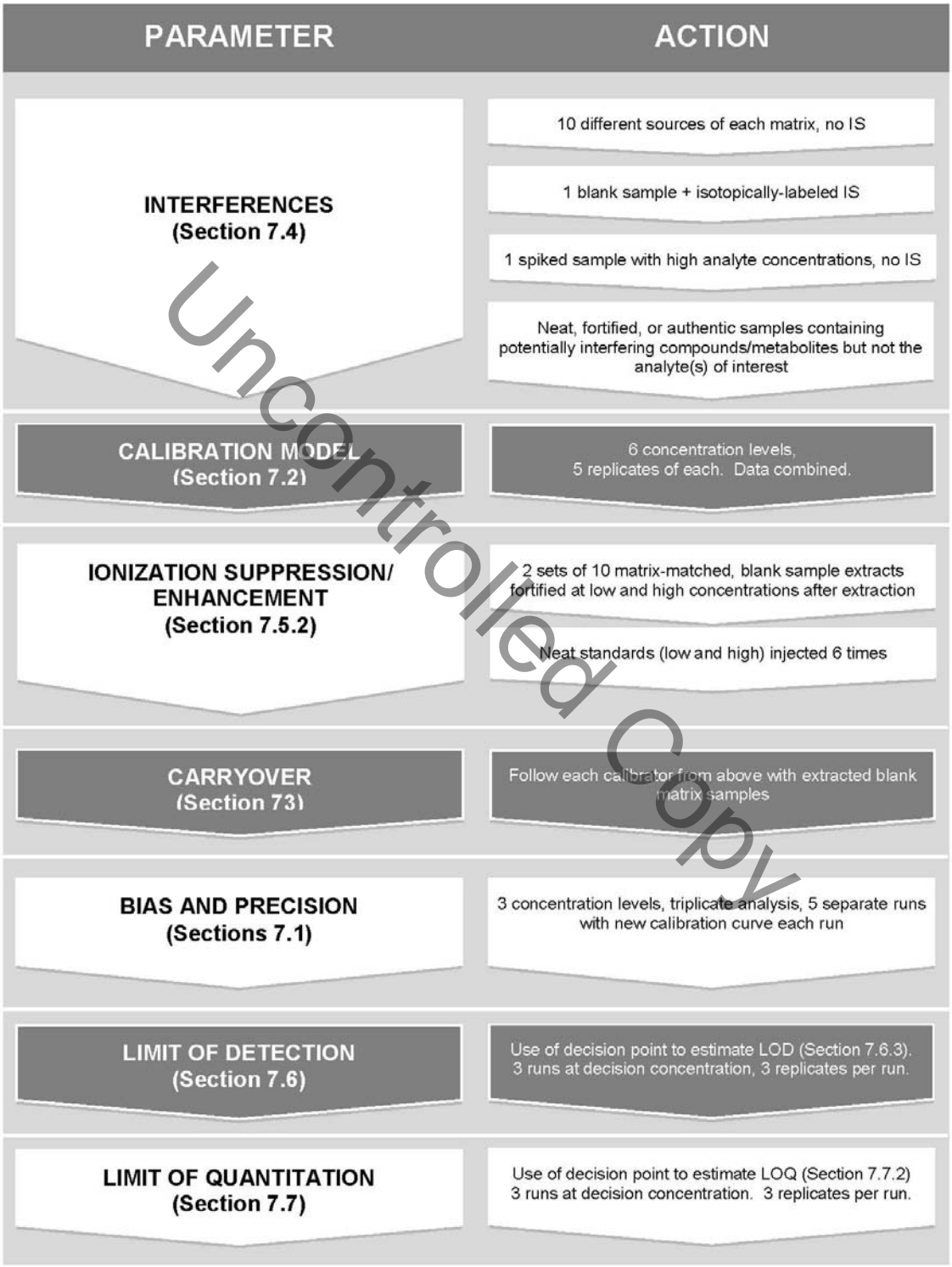
The laboratory used the decision point concentrations as the assay's limit of detection for each of the benzodiazepines and metabolites.

Documentation of Results (Section 10)
Along with all of the other required documentation listed in Section 10, the laboratory compared the results from the validation studies conducted to the originally defined requirements, as demonstrated in Table B3.

Table B3: Summary of validation results

Parameter:	Desired Limit:	Result:
Limit of Detection	Same as decision point:	
	Oxazepam (100 ng/mL)	100 ng/mL
	Lorazepam (100 ng/mL)	100 ng/mL
	Alprazolam (50 ng/mL)	50 ng/mL
	Alpha-hydroxyalprazolam (100 ng/mL)	100 ng/mL
Precision	% CV must not exceed 20%	Within-run (1.5 to 10.1%)
		Between-run (1.6 to 10.2%)

Appendix C: Example Flowchart of Method Validation Experiments



Appendix D: Table of Example Experiments for Validation of Qualitative Confirmation/Identification Methods

Interference (Section 7.4)

- 10 different sources of each matrix, no IS
- 1 blank sample + isotopically-labeled IS
- 1 fortified sample with high analyte concentrations, no IS
- Neat, fortified, or authentic samples containing potentially interfering compounds/metabolites but no analyte

Carryover (Section 7.3)

- Addressed in routine QC practices by analyzing extracted blank matrix samples between case samples

Limit of Detection (Section 7.6.4.1)

- Fortified matrix samples fortified at increasingly lower concentrations and analyzed in duplicate over 3 days. Lowest concentration that reproducibly yields signal greater than or equal to 3 times the noise of background signal is the LOD.

Appendix E: Table of Example Experiments for Validation of Quantitative Methods

Interferences	Ionization Suppression/Enhancement [a]	Calibration Model
• 10 different sources of each matrix, without IS • 1 blank sample with IS • 1 fortified sample with high analyte concentrations and without IS • Neat, fortified, or authentic samples containing potentially interfering compounds/metabolites but no analyte	**Post-column infusion:** • 10 blank extracts fortified after extraction at low concentration • 10 blank extracts fortified after extraction at high concentration • Analyte solutions for infusion (low and high concentrations) each injected 6 times	• 6 concentration levels, 5 replicates each (may be accomplished with calibration curves generated for studies below)

Main validation phase

		Bias & Precision					*Dilution Integrity*
Run	*Calibration*	*Low*	*Medium*	*High*	*LOD*[b]	*LOQ*[b]	*Bias & Precision*
1	6	3	3	3	3	3	3
2	6	3	3	3	3	3	3
3	6	3	3	3	3	3	3
4	6	3	3	3	-	-	3
5	6	3	3	3	-	-	3

[a] LC-MS(/MS) methods only
[b] For this example, the reference material approach is used to estimate the LOD (Section 7.6.4.1) and LOQ (Section 7.7.3)