P903: Policy on Estimating Measurement Uncertainty for ISO 15189 Testing Laboratories

TABLE OF CONTENTS

Introduction

A2LA has compiled information for classifying some common types of test methods according to this policy. This policy is intended as a means to facilitate transition to compliance with ISO 15189, and is subject to change as additional guidance is made available internationally.

This guidance has been developed and reviewed by the A2LA Medical Testing Technical Advisory Committee, provides information about how to categorize methods typically categorized. A2LA offers special thanks to Mr. Greg Cooper for his contributions to this policy. The medical laboratory must comply with 5.5.1.4 and 5.3.1.4 of ISO 15189 regardless of whether a method is listed below in Category 1, 2, 3 or 4. (See Section 3).

Measurement Uncertainty (UM) in medical laboratory testing is the doubt associated with what represents the trueness of a medical laboratory test result. The *International Vocabulary of metrology – Basic and general concepts and associated terms* (VIM) 3rd edition, definition 2.26 defines uncertainty of measurement as a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.

(Measurand is the quantity intended to be measured (VIM: 2012, definition 2.3), whereas analyte is an informal term used to identify the substance being measured.)

Indeed, virtually all test results are provided to a practitioner and/or patient without any indication to either party of the certainty of that measurement. For example, if today's glucose value is 110 mg/dl, and tomorrow's is 112 mg/dl, the question remains, "Are these results different?" On the surface, yes, these results appear different; in fact, one is considered a "normal" value, while the other is considered "abnormal." However, without understanding the variation inherent in the process used to measure glucose, such a judgment cannot be made. This variation contributes to the uncertainty of the measured value. Thus, understanding UM allows a practitioner to understand the probability of trueness of that value, subsequently leading to a better understanding of the clinical condition of that patient. Thus, to maximize the clinician's ability to effectively manage patient care, communication between the medical laboratory and the clinician regarding UM is desirable.

Medical Laboratories need Measurement Uncertainty.

Measurement uncertainty ultimately allows a practitioner better understanding of the clinical significance of a value, contributing to better clinical care of the patient.

- By understanding the amount of uncertainty in a measurement, a clinician can better understand the fitness of use of that measurement for clinical purposes.
- By controlling UM for applicable methods, laboratories can have assurance of the accuracy where accuracy represents both precision and trueness of their chosen methods.

By understanding UM, the clinician can understand the difference between normal process variation and variation that cannot be accounted for in the testing process. For example, UM does not account for

instances where a wrong patient was drawn and the subsequent analytical value is inconsistent with that patient's history.

Additionally, ISO 15189 (5.5.1.4) requires that:

"The laboratory shall determine measurement uncertainty for each measurement procedure in the examination phase to report measured quantity values on patients' samples.

There are 2 sources of variation that contribute to UM in the Medical Laboratory.

There is variation associated with the calibrators used in the medical laboratory methods.

- For example, a calibrator might indicate a value of 100 mg/dl; however, the manufacturer might state a level of uncertainty for that measurement.

- For Category 4 examinations only, if the manufacturer does not estimate uncertainty for its calibrator, the medical laboratory must do this by conducting another uncertainty of measurement for the calibrator.

There is also random variation associated with repetitive testing of the same sample in the test system.

- This common cause variation is also called "imprecision."

- These data are generally obtained from internal quality control data.

Refer to the Section "Identify the Components of UM" below for the management of these sources.

A number of contributors influence the value of the UM.

Thus the collection of data to estimate UM should include testing with different operators at different times with different pieces of equipment, for example.

2.0 General Terms Associated with Measurement Uncertainty

Accuracy
Closeness of agreement between a measured quantity value and a true quantity value of the measurand (VIM: 2012, definition 2.13)

Analyte
The substance measured from a sample. The characteristic of the substance that is measured is called the measurand, defined below.

Target Measurement Uncertainty
Measurement uncertainty specified as an upper limit and decided on the basis of the intended use of the measurement results (VIM: 2012, definition 2.34)

One widely used method to determine Target Uncertainty is to define the upper acceptable limit for imprecision as a proportion of the intra-individual biological variation of the analyte. This concept is sometimes referred to as analytical goal.

Bias
Estimate of a systematic measurement error. (VIM: 2012, definition 2.18)

Error
Difference between the measured quantity value of the measurand and the reference quantity value of the measurand. (VIM: 2012, 2.16)

Examinations
Clinical Tests performed in the Medical Laboratory.

Imprecision
Expressed variation, either standard deviation or coefficient of variation, calculated from the results in a set of replicate measurements.

Measurand
Quantity intended to be measured. (VIM: 2012, definition 2.3)

Metrological Traceability
Property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. (VIM: 2012, definition 2.41)

Numerical Significance
Figures of a number that have practical meaning. The number of significant figures used in a measurement expresses the degree of precision of the measuring system.

Precision
Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions. (VIM: 2012, definition 2.15)

Proficiency Testing
Inter-laboratory comparisons used to monitor performance of a medical laboratory with regard to individual tests, measurements, or observations. Also called “external quality assurance” (EQA) in many settings

Standard International Units (SI)
The system of metric units that is adopted by all major countries for use in science, medicine, industry, and commerce. It includes the base units as listed below:

- Length (m)

- Mass (kg)
- Time Interval (s)
- Electric Current (A)
- Thermodynamic Temperature (K)
- Luminous Intensity (cd)
- Amount of Substance (mol)

Measurement Uncertainty
Non-negative parameter, characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. (VIM: 2012, definition 2.26)

3.0 Critical Process for Estimating Measurement Uncertainty

There are essential actions that must be performed for a medical laboratory to successfully implement an Uncertainty of Measurement Program. These are found in the Table 1 below, and will be explained in the following sections.

Table 1:

Essential Actions	Output of Action
Categorize examinations	Document identifying examinations as Category 1, Category 2, Category 3, or Category 4. • Category 3 and Category 4 examinations will require UM identification. • Category 2 examinations will require evidence of compliance with the test method and compliance with required reporting instructions.
For each Category 3 or Category 4 examination, define the measurand of the method as well as clinically significant limitations and interferences.	Document identifying, for each Category 3 or Category 4 examination method, the measurand of the method along with any clinically significant limitations and interferences. • The identification of the clinically significant limitations and interferences must be from credible sources, and these sources should be known.
For each Category 3 or Category 4 examination, identify the components of UM.	Document identifying the UM components for each Category 3 or Category 4 examination. • Generally, this may be the long term QC imprecision data. However, it may also include the published UM of calibrators.

For each Category 3 or Category 4 examination, record means long-term imprecision QC data to serve as the estimate of UM.	Document identifying the UM for each Category 3 or Category 4 examination.
Where applicable, determine Target Uncertainty for each examination.	The reliability needed for a reported measured value may differ for each clinical purpose. For example, for what purpose is this analysis conducted and what would be the targeted UM such that the clinician can confidently apply the significance of the value for patient care?
Where applicable, compare and contrast the examination UM to the Target Uncertainty.	Document a review of significant contributors to total analytical imprecision when the Target Uncertainty is not met, as well as the resolution.
For each UM determined, identify the numerical significance to reflect the UM of the method.	Document the determination of numerical significance.
Coordinate with clinicians the availability of UM information.	Documentation of clinician consideration of necessity of UM information. Appropriate identification of UM parameter on test reports should be identified where determined to be necessary.
Monitor UM estimates over time.	Documentation of on-going monitoring, such as the collection and analysis of QC data over time.

Attention will now be given to each of these actions in detail.

4.0 Implementation of Action Steps

This section will provide a closer look at each action step so that the user can more easily implement a UM program.

Categorize Examinations

A medical laboratory should begin with the identification of all examinations within the scope of accreditation. The following categories should then be applied to determine which examinations warrant Uncertainty of Measurement determination.

1) *Category 1*: Methods that are reported on a qualitative basis, or on a categorical or nominal scale. In other words, there is an attribute answer to the examination inquiry. These methods are not kit driven. Here, UM is not meaningful because variation is due in large part to technical competency and not inherent in the measurement process itself.

- This category does not include methods that are reported on an ordinal scale, such as +, ++, +++, or ++++. It does include such an examination to determine the identification of a bacterial organism.

2) *Category 2*: *Well-recognized test methods that use kits or prepared reagents to determine qualitative results that are on an ordinal scale*. Here the lab satisfies the need for uncertainty estimates by following the published method, meeting performance requirements (such as proficiency tests) and reporting in accordance with the published method.

- Category 2 tests are those qualitative tests where all of the components of the analysis (of the kit) fall under a package insert.
    i. This package insert is approved by the FDA for human testing and version controlled by the manufacturer. Additionally, it contains uncertainty of measurement information appropriate for the test.

- Examples of this system would be the Rapid Plasmin Reagin, and Red Cell Antibody Identification Panels.

3) *Category 3*: *Well-recognized quantitative and semi-quantitative methods that are governed by FDA approved test systems, including package inserts for reagents and instrument manuals for equipment.*

- The medical laboratory must determine their uncertainty of measurement based on the reagent package insert performance criteria as well as the instrument performance criteria, and would typically use the internal quality control data to estimate uncertainty of measurement.
- Some test results provide a qualitative result (positive or negative) based on quantitative responses. If the cut-off point is pre-determined, the possibility of having "indeterminate" responses exists to account for the UM in the establishment of the cut-off. The results closest to the cut-off are most at risk, and should be used as the basis in determining the UM of the examination method.
    - In these cases, UM may be expressed as any one or more of the following:
        - Traditional UM statement for samples at levels near the cut-off value.
        - A statement about false classification rates for results near the cut-off.
        - Overall rates of correctness for different known classes of samples (known positives, known negatives, sensitivity, specificity, etc.)

4) *Category 4: In-house developed tests that generate quantitative data.*

- For these methods, UM must be estimated using available data, published information, and/or designed experiments.
- Some test results provide a qualitative result (positive or negative) based on quantitative responses. If the cut-off point is pre-determined, the possibility of having "indeterminate" responses exists to account for the UM in the establishment of the cut-off. The results

closest to the cut-off are most at risk, and should be used as the basis in determining the UM of the examination method.

- In these cases, UM may be expressed as any one or more of the following:
  - Traditional UM statement for samples at levels near the cut-off value.
  - A statement about false classification rates for results near the cut-off.
  - Overall rates of correctness for different known classes of samples (known positives, known negatives, sensitivity, specificity, etc.)

There should be some sort of evidence that this classification process took place for all examinations within the scope of accreditation. The laboratory should then prepare to provide the UM for all examinations requiring UM estimates. The following steps apply only to those examinations requiring the UM estimate.

Define the Measurand for Each Examination Requiring UM Estimates

For each Category 3 and 4 examination, the medical laboratory must specify what is actually being measured by the test method. For example, many examinations have high analytical specificity and thus measure only the substance they are designed to measure. However, some examinations are not as specific and thus tend toward cross reactivity and interfering substances which will be included in the final value. So, for each Category 3 and 4 examination, documentation must exist that identifies:

- The analytical method used.
- The substance the method is designed to measure.
- What is actually measured (the measurand).
- Any diagnostic limitations to the method.
- Any cross-reacting and interfering substances that impact the clinical interpretation of the values.
  - These should be known from credible and identifiable sources.

For example, when measuring for enzymatic activity, the examination does not necessarily directly measure the level of the enzyme, but rather the activity of the enzyme. So, in examinations for Alkaline Phosphatase, Alkaline Phosphastase activity is what is actually measured, not the Alk Phos level itself. Thus, Alk Phos is the analyte, and Alk Phos activity is the measurand.

Identify the Components of UM

As noted earlier, there are two sources of uncertainty in quantitative clinical laboratory testing that may exist. These include:

1) The normal variation or imprecision associated with running the method repetitively. So, if a known substance is assayed, its value may be, over time: 100, 105, 110, 102, 106, 108, 109, 107, 105, and 103. The standard deviation of these values can be termed the imprecision of the examination method. (Note: In no way does this example imply that 10 values should be used to calculate standard deviation; it is for illustrative purposes only).

   - This can be termed "common cause variation" because it is predictable in nature and inherent to the process. It cannot be eliminated, only reduced.

i. If the variation is not common cause, that is, it is not predictable in pattern, the process must be evaluated to bring it into a state of common cause variation (statistical control) before an imprecision measure can be made. In this case, assignable cause variation is present and must be eliminated.

2) The uncertainty associated with the calibrator used in the examination method.

- This UM should be provided by the commercial supplier.
- If the calibrator is not obtained from a commercial supplier, a UM estimate must be determined.

When both sources of variation exist, the UM of the examination system may consists of a total these sources. The key is to determine what the contribution of the components is and then decide what contribution is significant. In many cases, the calibrator will most likely result in a negligible contribution.

Record Mean Long-Term Imprecision QC Data as UM Estimate

For medical laboratory examinations, imprecision estimates such as standard deviation (SD) and coefficient of variation (CV) provide the UM estimate required, if the QC process includes all the steps and components involved in examining patient samples. These data are best collected over time across as many routine-operating conditions as possible to provide the most reliable estimate of UM. The SD or the CV represents the Quality Control data associated with the examination procedures.

- For established methods collect at least 6 months worth of internal QC data to calculate SD or CV
- For new methods use at least 30 data points for each level of QC using at least 2 different lots of calibrator and reagents, where applicable.
  - This provides a short term UM. Continue to evaluate until the long term can be established.
- For UM purposes UM represents the 95% confidence interval.
  - $\pm$2.0 SD
- In general, the more data collected, the more reliable the estimate.

NOTE: When medical laboratories test samples more than once and then use an "average" of the observations to report the final result, the repeatability (precision or imprecision in P603(e) contributor u(y)), to uncertainty may be reduced by the square root of the number (x) of repeated original results.

Establish the Target Uncertainty, Where Applicable; Compare UM to Goal

From BS ISO/TS 25680.7: "The value of the measurand coupled with the calculated UM should be compared to either a biological reference value or a clinical decision limit. Neither of these should have a UM."

The calculated UM should be evaluated to determine the significance of the measurand in impacting patient care. That is, if a patient result at the lower end of the uncertainty range would lead to a different clinical decision than a result at the upper end of the range, the uncertainty is too large.

Determine the Numerical Significance of UM Estimates and Clinical Results

The significant numbers used represent not only the value of a result, but also the certainty with which the result was determined. The key is to evaluate the imprecision data. Generally, the larger the imprecision value (SD or CV) the fewer significant digits that should be used. In other words, with great variation, 110 can look a lot like 107. But, with very little variation, 110 can be very different from 110.5. The medical laboratory must establish, based on its variation data, what represents a meaningful difference in values.

Interface with Clinicians on the Appropriate Use of UM Estimates

In many cases, UM estimates can contribute to patient care. Thus, in consult with clinicians, the medical laboratory must determine these results that could significantly impact clinical interpretation and subsequent patient management. It should also consider if it is necessary to include UM estimates in actual patient reports. As a result, it is expected that procedures exist in the medical laboratory for informing clinicians of significant UM information in a way that will be meaningful for clinical use. The expression of the UM concept is at the discretion of the medical laboratory such as a confidence expression. Now that there is an understanding of the activities necessary to establish UM estimates, attention will now turn to the monitoring of the UM.

5.0 Monitoring

As with any estimate, changes can occur with time. As more data are collected over time, the more stable the estimate becomes. Thus, for each examination with an UM estimate, an on-going monitoring program must be in place.

Components of a Monitoring Program

- Sampling Plan
    - The medical laboratory must determine when, where, and how QC data will be collected. Generally, all QC and PT data should be used to evaluate the established estimate of UM. The UM would then be adjusted based on these data.
    - These data are best collected over time across as many routine-operating conditions as possible to provide the most reliable estimate of UM.
    - The medical laboratory must determine how many measurements will be part of the sample. For example, perhaps the x-bar (mean) will be comprised of a certain amount of measurements, like 5 measures of one level of QC. Or perhaps each measurement will be tracked. Based on this sampling plan, the appropriate type of statistical chart can be used for monitoring, for example, an x-bar/R (average and range chart) or an xmR (individual measures and moving range chart).
- Rules of Monitoring
    - The medical laboratory must apply statistical rules to ensure that the examinations monitored are in a state of statistical control (exhibiting only common cause variation).

- When a process is exhibiting more than common cause variation, measures of imprecision, such as standard deviation and coefficient of variation are meaningless.
- Unacceptable results in proficiency tests should lead to a review of UM, unless a special cause source of variation is identified

- Actions to maintain currency of UM
  - The medical laboratory must define actions it will take when examinations with UM concern are not functioning within a state of statistical control.
  - This includes root cause analysis and subsequent corrective action.
  - This corrective action should consist of how the examination will be again monitored to ensure statistical control.
  - This action should also include how problems with the UM estimate will be communicated.
  - This action should also include how the new UM estimate will be re-established, documented, and communicated after statistical control is again achieved.
  - The medical laboratory must define how it will communicate and document changes in the estimate of UM when imprecision actually improves, with regard to a smaller standard deviation or coefficient of variation.

In conclusion, the medical laboratory has the responsibility to establish UM estimates for applicable examinations with regard to fitness for clinical use. As part of that establishment, the medical laboratory must implement a monitoring plan to ensure the integrity of it's UM estimates.

APPENDIX

Examples to Assist with Compliance with P903 Policy on Estimating Measurement Uncertainty for ISO 15189 Testing Laboratories

The calculation of measurement uncertainty (uncertainty of measurement) for analytical methods used in medical laboratory practice is not a simple matter. There has been much debate among medical laboratory professionals about the value of uncertainty and how true to the Guide to Uncertainty of Measurement (GUM) laboratories must be since this Guide was not originally intended for application to biological sciences. The Clinical and Laboratory Standards Institute (CLSI, Wayne PA published a consensus guidance on this subject matter in 2011 entitled ***C51 Expression of Measurement Uncertainty in Laboratory Medicine***. This guidance examines derivation of uncertainty for medical laboratory methods using GUM principles in a traditional approach and one that uses data already available in the laboratory; i.e. quality control data.

The following material is intended to examine the basics of uncertainty and to give some examples of "uncomplicating" its derivation. The CLSI document delves much deeper into the statistics of uncertainty and it is suggested that laboratories and assessors familiarize themselves with the content of this Guidance.

The first step in understanding measurement uncertainty is to understand why the authors of ISO 15189 thought it was necessary for medical laboratories to calculate and report this statistic. When the medical laboratory produces a test result such as 7.8 ng/mL for a serum prolactin, the value of 7.8 is NOT considered the TRUE concentration of prolactin. Rather, 7.8 is the medical laboratory's best estimate of the true value that, with a certain level of confidence, lies somewhere within an interval of values characterized by measurement uncertainty. Being able to provide the physician, on request or as needed, the measurement uncertainty associated with a particular test value helps the physician determine whether the patient is sick or well, the prognosis is good or bad, the treatment is effective or not effective. Regarding treatment effectiveness, the measurement uncertainty is particularly helpful in differentiating whether a change in a serial test result is due to analytical variation or a true physiological change.

A. When Is Calculation of Measurement Uncertainty (MU) Required By A2LA?

A2LA requires measurement uncertainty for any diagnostic test that produces a result that is expressed as a quantity. A2LA also requires measurement uncertainty for all calibrated ancillary equipment including but not limited to thermometers (including digital), balances, weights, centrifuges, timers, hygrometers and pipettes. This document will focus only on diagnostic tests and not on ancillary equipment, although the statistical principles described here also apply for ancillary equipment calibrations. Refer to the A2LA Policy *P905: A2LA Metrological Traceability Policy for ISO 15189 Laboratory Testing* for particular requirements for uncertainty for laboratory equipment**.**

A2LA divides diagnostic testing into the following four categories. Each category is associated with a specific requirement for Measurement Uncertainty as shown in Table 2.

Table 2: Summary of Measurement Uncertainty Calculation Requirements Based on Category of Testing

Category	Classification	Description	UM requirement	Example
Category 1	FDA-cleared or approved or Lab Developed Test (LDT) (validated)	Results reported as Pos/Neg (Qualitative)	No MU required.	RPR agglutinin test
Category 2	FDA-cleared or approved	Results reported as +1,+2, +3 etc. (Qualitative)	No MU required.	Urobilinogen on a Urinalysis strip
Category 3	FDA-cleared or approved	Results reported as a measured quantity (Quantitative)	MU estimate using long-term imprecision (s) with adjustment for bias	Any test on a device that gives a direct read out as a quantity. Qualitative tests based on a quantifiable absorbance cutoff.
Category 4	Modified FDA-or approved Test or LDT or Lab Modified Test	Results reported as a measured quantity (Quantitative)	MU estimate using long-term imprecision (s) with adjustment for bias. When this is not possible, MU estimate based on and consistent with Guide to Uncertainty of Measurement (GUM)	Any test developed by the laboratory that expresses the result as a quantity. Any test that is modified by the laboratory that expresses the result as a quantity.

B. Uncertainty of Measurement: Overview of the Statistics

When reporting tests falling in Category 1 or 2, the estimation of measurement uncertainty is not required. See Table 2 above.

When measurement uncertainty is reported for a Category 3 or 4 test, A2LA requires the measurement be reported as an expanded uncertainty (U). This statistic is derived from two statistical precursors, standard uncertainty (u) and combined uncertainty (u_c).

Standard uncertainty (u) is expressed as a standard deviation [s] for a set of n replicates in a sample (y). This is expressed as:

$$u(y) = s(y)$$

where:
u(y) is the standard uncertainty of a data set (y) with n values
s(y) is the standard deviation of n values in data set (y)

Combined uncertainty (u_c) accounts for multiple sources of variability within a process such as might occur with measurement of the quantity of a substance in a human sample each source having its own uncertainty. For A2LA purposes, the standard deviation of a set of values derived from testing with an FDA-cleared or approved method/kit/device (Category 3 tests), previously defined as u_y, already accounts for these variables. So it follows then that for A2LA purposes:

$$s(y) = u(y) = u_c(y)$$

But for clinical diagnostic testing, this statistic underestimates the uncertainty because it does not account for any bias that may be present in the diagnostic method. So for A2LA purposes, and in keeping with the spirit and intent of CLSI's C51 guidance, the formula is modified as follows:

$$u_c(y) = \sqrt{(u(y)^2 + u_{Bias}(y)^2)}$$

where:
√ is the square root
$(u_{(y)}^2 + u_{Bias\,(y)}^2)$ is the sum of the squares for the standard uncertainty of the data set and the standard uncertainty of the bias for the data set.

A new statistical term was just introduced; that being $u_{Bias\,(y)}$. This statistic is the standard uncertainty associated with any bias in the testing process. It requires knowledge of:

- the individual data points in the set of test data,
- the true value and associated standard uncertainty of a reference material if a reference material is used to estimate bias
- the long-term mean and standard deviation of a peer group from an interlaboratory comparison program if a peer group is used as the reference to establish bias
- the number (n) of data points in the peer group
- the bias for each data point generated

This information is used to calculate $u_{Bias}(y)$ which is defined as:

$$u_{Bias}(y) = \sqrt{(RMS_{Bias}(y)^2 + u_{Ref}^2)}$$

where:

$RMS_{Bias}(y) = \sqrt{(\Sigma\ Bias(y)^2/n)}$

To calculate Bias you need to know either the true value of a reference material or the long-term mean of a peer group.

u_{Ref} = the uncertainty reported by the manufacturer for the reference material or if a peer group is used, it is the standard deviation of the data set divided by the square root of the number of values used to calculate the standard deviation ($s/\sqrt{n}$).

A2LA requires that measurement uncertainty be reported in the form of an expanded uncertainty (U). To calculate the expanded uncertainty (U) for an FDA-cleared or approved method/kit/device (Category 3 tests) or for lab developed tests (LDTs), modified FDA cleared or approved tests or in-house calibrated equipment (Category 4 tests), the expanded uncertainty (U) is calculated as:

$$U(y) = k * u_c(y) = k*\sqrt{(u(y)^2 + u_{Bias}(y)^2)}$$

where:

U(y) = the expanded uncertainty
k = a constant for the coverage interval containing the set of true quantity values of the measurand with a stated probability. For a 95 % probability the coverage factor used is 2
$u_c(y)$ = the combined uncertainty for a data set (y) with an adjustment for bias

To comply with A2LA Policy P903, the report of uncertainty must be an expanded uncertainty (U) and include the quantity (i.e. the measure), the units, the k factor of 2 for the coverage interval, and a statement of the level of confidence (95% level of confidence) for the calculation. For example: The expanded uncertainty (U) for pH of a calcium buffer is 0.001 units using a k factor of 2 with a confidence interval of 95%.

If the test was developed (LDT) or modified by the laboratory (Category 4 tests), the estimate of measurement uncertainty must use long-term imprecision (s) with adjustment for bias. When this is not possible, the laboratory must follow the Guide to Uncertainty of Measurement (GUM) or the CLSI C51 guidance on uncertainty using the bottom up approach; i.e. each variable would have to be identified and accounted for in the estimate of combined uncertainty as shown in the following formula:

$$u_c(y) = \sqrt{(u_{(y)}(1)^2 + u_{(y)}(2)^2 ..... + u_{(y)}(N)^2)}$$

where the combined uncertainty for data set (y) is the square root of the sum of the squares of the standard uncertainty for <u>each contributor</u> to variability in the overall process.

C. Measurement Uncertainty for Category 3 Tests

A2LA Policy P903 allows medical laboratories to use long-term imprecision (standard deviation) for estimating the uncertainty of measurement for unmodified FDA-cleared or approved tests. However to be consistent with the CLSI guidance on measurement uncertainty (C51), the uncertainty estimate must also account for analytical bias. The following describes a stepwise process to calculate uncertainty.

PROLACTIN EXAMPLE:

We wish to know the uncertainty for serum prolactin. The laboratory uses a floor model analyzer purchased from a leading diagnostic manufacturer. The analytical method used is chemiluminescence and results are reported in ng/mL. QC data collected in the month of November 2011 on a normal and high concentration control material are used to estimate the standard uncertainty of the data collected. Bias estimation and calculating the uncertainty of the bias is based on cumulative peer group data (mean, standard deviation, number of points) obtained from the interlaboratory comparison program representing greater than 6 months of data.

TO CALCULATE THE EXPANDED UNCERTAINTY (U) FOR EACH LEVEL (CONCENTRATION) OF CONTROL MATERIAL, DO THE FOLLOWING:

1. Collect QC data
   - Good: Minimum of one month or 30 data points
   - Better: 60-90 data points or 2-3 months of data
   - Best: 180 data points or 6 months of data

NOTE: For the following example, the samples were tested once and final results were not "averages" of replicate observations. QC data points were then used to calculate the standard deviation to estimate the measurement uncertainty at a particular concentration as determined by the concentration of the control material; these data points were individual observations obtained from 47 separate analytical runs. (See step 2.)

2. Calculate the standard deviation (s) of the data set chosen (y)
   - The standard uncertainty (u) for the data set is equal to the standard deviation [u(y) = s(y)]

3. Identify the following from interlaboratory comparison reports
   - Peer group cumulative mean
   - Peer group cumulative standard deviation
   - The number of points (n) in the peer group

4. Calculate the standard uncertainty of the peer group [$u_{Group} = s/\sqrt{n}$]
   - Divide the reported standard deviation of the peer group by the square root of the number of data points in the peer group used to calculate the reported standard deviation.

5. Calculate the root mean square (RMS) [$RMS_{Bias} = \sqrt{(\Sigma\ Bias^2/n)}$]

   - Calculate the bias for each observation (value) in the data set
     - Use the appropriate peer group mean value as the target value
       - Defined as same instrument, analytical method , reagent manufacturer, units of measure and temperature where appropriate.
   - Square each bias value and sum all the squares
   - Divide the sum of the squares by the number (n) of values in the data set
   - Take the square root of the result

6. Calculate the standard uncertainty of the bias [$u_{Bias} = \sqrt{(RMS_{Bias}^2 + u_{Group}^2)}$]

   - Square the RMS statistic and the standard uncertainty of the peer group. Sum these two squares and take the square root of the sum.

7. Calculate the combined uncertainty of the data set.

   - Combine the standard uncertainty of the QC data set with the standard uncertainty of the bias for the data set by squaring the standard uncertainty of the data set and the standard uncertainty of the bias for the data set. Sum these two squares and take the square root of the sum. [$u_C = \sqrt{(u(y)^2 + u_{Bias}^2)}$]

8. Calculate the expanded uncertainty (U) for a 95% confidence interval.

   - Multiply the combined uncertainty calculated above and multiply by a factor of 2.

Test	Prolactin			Reviewed by:			
Instrument	XYZ			Approved by:			
Method	Chemiluminescence			Date:			
Units	ng/mL						
Data for:	Nov. 2011						
		Level 1	x_{obs} (L1)	Bias	Level 3	x_{obs} (L3)	Bias
Mean reported for the peer group of 417 individual observations (data points)	Group Mean	7.5			35		
Number of individual observations for the uncertainty estimate (data points)	n_{obs}=	47			47		
Number of individual observations (data points) reported for the peer group	n_{Group} =	417			398		
Standard deviation reported for the peer group based on 417 points	s_{Group} =	0.512			1.68		
			8.3	0.8		37.7	2.7
			8.3	0.8		37.8	2.8
			7.2	-0.3		34.1	-0.9
			7.2	-0.3		34.8	-0.2
			7.3	-0.2		34.7	-0.3
			7.2	-0.3		34.4	-0.6
			7.1	-0.4		34.5	-0.5
			7.1	-0.4		34.3	-0.7
			7.2	-0.3		35.5	0.5
			7.2	-0.3		35.3	0.3
			7.3	-0.2		35.1	0.1
			7.6	0.1		35.4	0.4
			7.5	0		35.8	0.8
			7.4	-0.1		35.3	0.3
			7.5	0		36.2	1.2
			7.5	0		35.7	0.7
			7.3	-0.2		34.9	-0.1
			7.4	-0.1		35.2	0.2
			7.1	-0.4		35.3	0.3
			7.3	-0.2		35.2	0.2
			7.2	-0.3		33.9	-1.1
			7.4	-0.1		34.7	-0.3
			7.4	-0.1		34.8	-0.2
			7.4	-0.1		35.3	0.3

			7.3	-0.2		36.1	1.1
			7.5	0		37.1	2.1
			7.2	-0.3		35.3	0.3
			7.1	-0.4		35.8	0.8
			7.5	0		35.4	0.4
			7.9	0.4		37.3	2.3
			7.8	0.3		36.7	1.7
			7.9	0.4		37.4	2.4
			7.8	0.3		36.4	1.4
			7.9	0.4		35.5	0.5
			7.8	0.3		34.8	-0.2
			7.7	0.2		35.5	0.5
			7.6	0.1		35.6	0.6
			7.9	0.4		37.1	2.1
			7.9	0.4		36	1
			7.5	0		35.4	0.4
			7.6	0.1		37	2
			7.9	0.4		36.5	1.5
			7.5	0		35.9	0.9
			7.7	0.2		35.4	0.4
			7.6	0.1		36.1	1.1
			7.9	0.4		36	1
			8	0.5		35.7	0.7
$RMS_{Bias} = \sqrt{(\Sigma\ Bias^2/n)}$				0.31			1.14
$u_{obs} = sd(x_{obs})$				0.31			0.94
$u_{Group} = s/\sqrt{n}$				0.03			0.08
$u_{Bias} = \sqrt{(RMS_{Bias}^2 + u_{Group}^2)}$				0.31			1.14
$u_C = \sqrt{(u_{obs}^2 + u_{Bias}^2)}$				0.44			1.48
$U = k^* u_C$ (95 %) with a k of 2				0.88			2.96

Returning to the discussion of the prolactin result of 7.8 ng/mL at the beginning of this appendix, we can say with 95% confidence that the TRUE value for the prolactin result lies within the interval of 6.92 ng/mL and 8.68 ng/mL (7.8 ± (2*0.44). To meet A2LA reporting requirements for uncertainty, the medical laboratory would report the expanded uncertainty (U) for the normal concentration of prolactin as 0.89 ng/mL with a 95% confidence interval using a k factor of 2.

D. Measurement Uncertainty for Category 4 Tests (Lab Developed Tests and Modified Tests)

For most Category 4 examinations, the approach described for Category 3 can apply.

When this is not possible, measurement uncertainty for Lab Developed Tests (LDTs) and Modified Tests requires the medical laboratory to undertake a measurement uncertainty assessment consistent with principles contained in the Guide to Uncertainty of Measurement (GUM). Methods to establish measurement uncertainty involve one of two approaches: replication experiments (Type A) and/or information from current literature, available or pre-existing laboratory documentation, research or manuals/product inserts/printed materials supplied by the manufacturer (Type B). Medical laboratories may also refer to the calculation models described in the Clinical and Laboratory and Standards Institute (CLSI) guidance on measurement uncertainty; C51 Expression of Measurement Uncertainty in Laboratory Medicine.

In general, measurement uncertainty assessments according to the GUM model must account for as many sources of variation (uncertainty) associated with the measurement as possible including pre-measurement and post-measurement sources as well as those sources associated with the measurement itself. CLSI guidance on measurement uncertainty states that "only uncertainty sources pertaining to the measurement should be included in the measurement uncertainty. However, some or all of the pre- and post measurement sources generally have an effect on the reported result and, therefore, potentially on how it is interpreted by the user." …Based on this statement, the laboratory must at least give some consideration to pre- and post-measurement sources when creating the uncertainty budget for the test. Some examples of pre-measurement sources of uncertainty in the medical laboratory may include but are not limited to:

1. Patient preparation
2. Specimen collection
3. Volume of anticoagulant in the collection tube
4. Time lapse between specimen collection and measurement
5. Centrifugation of the sample (time and RPM)
6. Transportation and storage of the sample
7. Manual and automated aliquoting of samples (Note 1)
8. Staff (test operator) competency particularly for technique sensitive procedures

Note 3*: Manual methods are particularly susceptible to operator technique and tolerances for automated methods may erode over time.*

However, the CLSI guidance goes on to assume that quality systems in place in the medical laboratory and the institution at large often mitigate the contribution of pre- and post-measurement sources of variation in the medical laboratory. Again quoting C51, "Pre- and post-measurement uncertainties may be difficult to estimate and treat correctly. In laboratory medicine, it is common practice to minimize—where possible—the pre- and post-measurement uncertainties by implementing standardized procedures… [and] it is assumed that measurements are conducted according to the relevant procedure and without blunders or other technical non compliances."

For the purpose of A2LA accreditation and compliance with the A2LA Policy P903 on uncertainty, pre- and post-measurement sources of uncertainty will not be required for the uncertainty assessment as long as the

medical laboratory can provide evidence these contributors were considered and the medical laboratory can provide evidence that these sources were determined by the laboratory management to have no or insignificant effect on the test result.

E. Calculation of Expanded Uncertainty for Category 4 Tests: Lab Developed Tests (LDTs)

A measurement uncertainty assessment for LDTs can be and often is more substantial than for modified tests, because much of the information critical to a reliable assessment must be identified and developed by the laboratory itself. The medical laboratory should begin by preparing a process map identifying points along the process where variability might occur. A cause and effect diagram like a fishbone/Ishikawa diagram helps to further characterize contributors to variability; i.e. potential uncertainty components. The outcome of this work is an uncertainty budget that should include all contributors to uncertainty in the actual measurement procedure.

The medical laboratory must identify the contribution of all known pre- and post-measurement sources of variability and determine whether these sources must be accounted for in the uncertainty budget.

An example of an uncertainty budget for an LDT is provided in Table 3 and may include but not be limited to the components listed.

Table 3: An Example of an Uncertainty Budget[1]

Component	Standard Uncertainty[2] u(y)	Type[3,4]
Any significant pre- or post-measurement contributor (x), listed individually	u (x)	Type A (replication) or Type B
Sample indication (measurement signal)	u (S_s)	Type A (replication)
Calibration indication (measurement signal)	u (S_{cal})	Type A (replication)
Blank indication (measurement signal)	u (S_o)	Type B
Calibration	u (C_{cal})	Type B(manufacturer)
Dilution factor, as appropriate	u (d)	Type B
Matrix effects (due to sample matrix)	u (ME_s)	Type A (replication) or Type B
Matrix effects (due to interfering substances – list each individually)	u (ME-IS_n)	Type A (replication) or Type B
Sensitivity (Limits of detection)	u (LOD)	Type A (replication)
Specificity	u (SP)	Type A (replication)
Biological Variation of the measurand	u (BV)	Type A (replication) or Type B
Pipette calibration if calibrated pipettes are used	u (P	Type A (replication)
Other sources (e.g. Technique for technique sensitive tests)	u (O)	Type A (replication)

[1]Table modified from the Table used in CLSI C51 Expression of Measurement Uncertainty in Laboratory Medicine

[2]Uncertainties in the budget are often standard uncertainties (standard deviations) but may be combined uncertainties as well. One example in this Table of a component that would be a combined uncertainty would be the uncertainty attributable to the calibration.

[3] Please note that standard uncertainties (u_y) may be either Type A (replication) or Type B (information) or Both. Type B uncertainties expressed as expanded uncertainties (U) will have to be converted to a standard or combined uncertainty

[4] Type B uncertainties are often expressed as expanded uncertainties (U) that need to be converted to standard uncertainties. For example, a manufacturer reports the concentration of a key reagent for the test as X ± 2% at a level of confidence of 95%, k=2. This type of expression is known as expanded uncertainty (U) because it is accompanied by a coverage interval and is defined as:
$U = k * u_c(y)$

Solving the equation, the combined uncertainty $u_c(y)$ of the example given can be determined as U/k or 2 %/2 = 1 % or .01 as a ratio.

If the Type B uncertainty is reported as an interval of possible results (e.g. 1.2-3.4 mmol/L) , the standard uncertainty u(y) = a/ √3 where a = half of the assumed interval. Solving the equation for this example gives a standard uncertainty of 2.4/√3 mmol/L or 1.3 mmol/L.

Once all the uncertainties are determined for each identified source of measurement uncertainty in the budget, a combined uncertainty (u_c) for the LDT example is calculated as follows:

$$u_c = \sqrt{u(x)^2 + u(S_s)^2 + u(S_{cal})^2 + u(S_0)^2 + u(C_{cal})^2 + \ldots + u(P)^2 + u(O)^2}$$

To complete the exercise in compliance with A2LA Policy P903, the uncertainty is reported as an expanded uncertainty (U) described earlier along with a statement of the level of confidence and k factor used.

Table 4: Example of Calculating an Expanded Uncertainty for a Lab Developed Test

Component	Standard Uncertainty u(y)	Type[1]	Comment
Any significant pre or post measurement contributor (x), listed individually	u (x)= 0 mmol/L		No notable contributors found. Records of examination of pre and post contributors can be found in the Quality Manager's Office for the Research Department.
Sample indication (measurement signal)	u (S_s) = .05 mmol/L	Type A (replication)	Replicate experiment performed on 150 serum samples (serum samples from 5 different individuals every day for 30 days)
Calibration indication (measurement signal)	u (S_{cal}) = .002 mmol/L	Type A (replication)	Replicate experiment with 5 lots of serum calibrator at three concentrations other than zero. ("G" mmol/L, "H" mmol/L and "J" mmol/L
Blank indication (measurement signal)	u (S_o) = 0	Type B or Type A (replication)	Replicate experiment with 5 serum blanks performed each day for 30 days totaling 150 samples.
Calibration	u (C_{cal}) = 0.005 mmol/L	Type B (manufacturer) or Type A (replication)	The u_c was determined in accordance with ISO 17511
Dilution factor, as appropriate	u (d) = 0 mmol/L	Type B	Not applicable
Matrix effects (due to sample matrix)	u (ME_s) = 0 mmol/L	Type A (replication) or Type B	Matrix effects for unadulterated human serum are disregarded
Matrix effects (due to interfering substances – list each individually)	u (ME-IS_n) = 0 mmol/L	Type A (replication) or Type B	This test was performed on treated with "abc" drug and no interferences were found. Hemolysis

			interferes with the measurement so hemolyzed specimens must be rejected.
Sensitivity	u (LOD) = 0.15 mmol/L	Type A (replication)	The ability of the method to distinguish measurand from zero is 98.1 %
Specificity	u (SP) = .06 mmol/L	Type A (replication)	The ability of the method to differentiate the measurand from other epitopes is determined to be 99.1%
Biological Variation of the measurand	u (BV) = 1.2 mmol/L	Type A (replication) or Type B	Determined though replicate testing of 400 samples of serum from a mixed age and gender population.
Pipette calibration if calibrated pipettes are used	u (P) = 0 mmol/L	Type A (replication)	Not applicable
Other sources (e.g. Technique for technique sensitive tests)	u (O) = 0 mmol/L	Type A (replication)	Not applicable

Calculate the expanded uncertainty for LDT (XYZ).

$u_c = \sqrt{(u(x)^2+u(S_s)^2+u(S_{cal})^2+u(S_o)^2+u(C_{cal})^2+u(d)^2+u(ME_s)^2+u(ME\text{-}IS_n)^2+u(LOD)^2+u(SP)^2+u(BV)^2+u(P)^2+u(O)^2)}$

$u_c = \sqrt{((0)^2+(0.05)^2+(0.002)^2+(0)^2+(0.005)^2+(0)^2+(0)^2+(0)^2+(0.15)^2+(0.06)^2+(1.2)^2+(0)^2+(0)^2)}$

$u_c = \sqrt{((0)+(0.0025)+(0.000004)+(0)+(0)+(0.000025)+(0)+(0)+(0)+(0.0036)+(1.44)+(0)+(0))}$

$u_c = \sqrt{1.446129}$

$u_c = \pm 1.2$ mmol/L

$U = k * u_c$

For 95 % level of confidence k = 2

$U = \pm 2*1.2$ mmol/L

$U = \pm 2.4$ mmol/L (k=2)

The expanded uncertainty (U) for test LDT (XYZ) is 2.4 mmol/L at a 95 % level of confidence

F. Calculation of Expanded Uncertainty for Category 4 Tests: Modified FDA-Cleared or Approved Tests

As with LDTs, modified tests require an uncertainty budget. Modifications may take two forms.

1. The medical laboratory modifies its own previously developed LDT.
2. The medical laboratory modifies a test certified by an external source, such as the FDA

(1) Modifying a previously developed LDT

The medical laboratory merely needs to revise its previously developed uncertainty budget for only those items that are affected by the modification and recalculate the combined uncertainty.

(2) Modifying an externally certified test

When an externally certified test is modified, the medical laboratory must determine whether the modification:

1. requires development of an uncertainty budget leading to recalculation of the measurement uncertainty or
2. using the long term standard deviation of the modified test accounting for the uncertainty of the bias in the established measurement.

Note 6: *Should the medical laboratory undertake option 1, many of the components of uncertainty for the unmodified test required for the budget may be available from the manufacturer. If not, the laboratory will have to estimate the uncertainties for those components not affected by the modification. Those components that are affected by the modification will have to be determined by the medical laboratory and inserted into the budget as revised estimates.*

References

1) BS ISO/TS 25680: Medical Laboratories – Calculation and expression of measurement uncertainty. British Standards Institute (BSI), 2013.

2) ISO 15189 (2012): Medical Laboratories – Requirements for quality and competence. ISO, 2012.

3) International vocabulary of metrology – Basic and general concepts and associated terms (VIM). JCGM 200:2012, 3rd edition.

4) C51: Expression of Measurement Uncertainty in Laboratory Medicine. The Clinical and Laboratory Standards Institute (CLSI), 2011.

Document Revision History

Date	Description
2009	➢ First issued
April 2013	➢ Revised to address ISO 15189 (2012) ➢ Revised to address VIM: 2012 ➢ Revised to add the Appendix ➢ On page 18, clarified that this calculation based on a single data set ➢ On page 11, added language to clarify how a lab would handle averaging the data set, if it were to occur ➢ On page 17 and 26, added parentheses to the equations ➢ On page 20, added definitions for nGroup, sGroup and uGroup to the Table ➢ On page 21, corrected the values in the expanded uncertainty equation example ➢ Added Reference section ➢ On page 4, changed "uncertainty" to "error" ➢ On page 4, Section 2.0, changed "Terms" to "General Terms" ➢ On page 6, Section 3.0, changed "table below" to "Table 1 below" ➢ Updated A2LA Header
February 2014	➢ On page 4, the term "Error" replaced with "variation" and better clarified ➢ In Table 2, the requirements for Category 1 and 4 tests were clarified. ➢ On page 5, "Inaccuracy" term removed
3/25/14	➢ Updated document titles/references throughout. Key elements made consistent with P703.