COLLEGE of AMERICAN PATHOLOGISTS

Measurement Uncertainty Guide

ISO 15189 Accreditation Program

(2026 Update)

Authors
Rohit Gulati, MD FCAP
Rich Becker, MSEd, CQIA

Reviewers
Deborah Miller-Jones, MLS(ASCP), MPH
Jeremy Hart, MD FCAP MBA

Table of Contents

This page intentionally left blank

Preface to the Guide

This guide serves two audiences:

- Those who want clarity on ISO 15189 requirements and assurance that their laboratory practices conform.

- Those who want a deeper understanding of measurement uncertainty (MU).

    - How to calculate and express MU for different domains – within and outside clinical pathology
    - How MU serves risk management
    - How to communicate MU with clinicians – sample consultations

The basic information on requirements can be found in the first section on Intent, Requirements, and Implementation.

Our advice is to read these pages, then skim through the rest in case it is needed.

Note on scope of MU:

While MU has historically been defined in a purely quantitative manner, the ISO 15189:2022 update prompts users to identify sources of uncertainty in a qualitative manner, going beyond the measurement system.

In this expanded definition, MU refers to:

- Specific and calculable uncertainty of the measurement system, expressed in SD or related concepts like CV
- Uncertainty of a test or result from sources that may not always be quantifiable, including qualitative uncertainty of the test or result, taking into account the following:
    - Uncertainty due to biological variation
    - Uncertainty due pre-examination phase factors

Intent, Requirements, and Implementation

Intent of the ISO 15189:2022 measurement uncertainty (MU) clause

The MU clause seeks to help laboratories enhance their understanding of:

- The sources of MU
- How to use it to make clinical decisions and improve processes

Specifically, clause 7.3.4 seeks to make sure the following are true for the laboratory:

1. **The laboratory understands how "reliable" and "trustworthy" its test results are.**

   Measurement uncertainty (MU) is basically the "range of possible error" around a laboratory result.

   The standard intends for laboratories to know how close their measured value is to the true value.

   Think of it like saying:
   "My ruler is good, but it could be off by about ±1 millimeter."

   Laboratories must know this for their tests too.

   Most high-quality laboratory practices have measurement uncertainty ingrained in their test validation/ verification/ operations/ result interpretation and most clinical best practice guidelines incorporate uncertainty in clinical implication that follow laboratory test results.

2. **The laboratory can explain how much confidence one should have in a test result.**

   MU helps doctors and patients understand how sure the laboratory is about a number, or a result, especially for things like glucose, cholesterol, hemoglobin, etc.

   It's not saying the test is wrong.

   It's saying:
   "Here is the number, and here's how sure we are about it."

   OR
   "The blood glucose is 124 mg/dL; and we are 95% confident that patients' actual blood glucose value is within 118 - 130 mg/dL (CV~5%) of the reported value."

   OR
   Test result: Fasting Blood Glucose: 124 mg/dl.

   > Instead of saying: "The glucose result is accurate." Since all QC results are acceptable
   >
   > In essence, we may now say:
   >
   > "Given all details about this measurement process, we are 95% confident that patients' actual blood glucose value is within 118 - 130 mg/dL (CV~5%)"

   Note: Clinical providers and laboratory clients may not be well versed in using the most specific terminology such as MU, coefficient of variation, or confidence intervals. More likely, they will ask, "How "accurate" or "reliable" is the result?"

3. **The laboratory uses MU in meaningful places—not everywhere.**

   ISO 15189 does not require MU calculations for every step in the testing process.

   The intent is for laboratories to apply MU to key results where uncertainty really matters for medical decisions.

This brings us to risk identification/ management in the laboratory's role in clinical decision support.

Examples:
- borderline results (possibly repeat testing is valuable)
- results used for diagnosis
- results with strict clinical cutoffs
- results used for monitoring change over time

4. **The laboratory uses MU to continually improve processes and <u>make better decisions.</u>**

MU should help the laboratory decide things like:
- Is the method performing well enough?
- Are two methods close enough to replace each other?
- Is the test precise enough to detect small changes in a patient over time?
- Are QC results acceptable?

The intent is continuous improvement, not paperwork.

5. **The laboratory should be able to show its work if asked.**

Laboratories need to know:
- where their MU numbers come from,
- the data they used,
- and how they keep MU up to date.

This shows the laboratory understands its system and can defend (explain) its numbers (results).

In summary, to make sure the laboratory:
1. Knows the limits of its testing system.
2. Knows how confident it can be in a number.
3. Uses MU only where it matters/ its relevant clinically.
4. Uses MU to improve quality.
5. Can explain MU if asked.

Why was the term MU introduced?

Measurement Uncertainty (MU) comes from metrology (the science of measurement) and was introduced to improve how we describe the reliability of laboratory test results. Specifically:

1. To replace confusing or misleading terms
Several terms were historically used to describe the quality of test results, but each had limitations:
- **"Error"** (difference between measured and true value)
  - Implies a mistake or a faulty test, which can be misleading for clinical teams when the test result is actually valid.
- **"Accuracy"** (closeness to true value)
  - Often used inconsistently and can mean different things to different people, leading to confusion.
- **"Precision"** (variation between individual measurements performed by replicate testing)
  - Describes repeatability (how close repeated results are to each other), but not correctness

MU was introduced to avoid these misunderstandings and provide a clearer, more consistent concept.

2. Focus on "range" and "probability" and to acknowledge that "true" value cannot be identified.

3. To reinforce that all measurements are estimates
The concept of MU emphasizes that laboratory results are **estimates**, not exact truths. It helps account for:
- **Random variation** (imprecision, analytical noise)
- **Systematic effects** (bias, calibration limitations)

- **Environmental and biological variability**

Together, these factors influence every reported test result.

Requirements – What does ISO 15189:2022 require regarding MU?

Clause 7.3.4 states as follows:

7.3.4 Evaluation of measurement uncertainty (MU)

a) The MU of measured quantity values shall be evaluated and maintained for its intended use, where relevant. The MU shall be compared against performance specifications and documented.
NOTE ISO/TS 20914 provides details on these activities together with examples.

b) MU evaluations shall be regularly reviewed.

c) For examination procedures where evaluation of MU is not possible or relevant, the rationale for exclusion from MU estimation shall be documented.

d) MU information shall be made available to laboratory users on request.

e) When users have inquiries on MU, the laboratory's response shall take into account other sources of uncertainty, such as, but not limited to biological variation.

f) If the qualitative result of an examination relies on a test which produces quantitative output data and is specified as positive or negative, based on a threshold, MU in the output quantity shall be estimated using representative positive and negative samples.

g) For examinations with qualitative results, MU in intermediate measurement steps or IQC results which produce quantitative data should also be considered for key (high risk) parts of the process.

h) MU should be taken into consideration when performing verification or validation of a method, when relevant.

In summary:

1. Maintain MU values.
2. Compare MU against performance specifications.
3. Review MU evaluations regularly.
4. Document the rationale for exclusions when MU is not possible or relevant.
5. Make MU available to laboratory users upon request.
6. In answering inquiries, take into account sources of uncertainty beyond measurement uncertainty, such as biological variation or environmental/pre-examination factors.
7. Estimate MU for positive negative tests based on representative positive and negative samples.
8. For qualitative examinations, include MU from intermediate quantitative steps or IQC data in high-risk parts of the process. Monitor MU in the intermediate quantitative step of high significance.
9. When relevant, consider MU during method verification or validation.

For a comparison of the requirements in the 2012 and 2022 versions of ISO 15189, see Appendix E.

Implementation – CAP Guidance Summary

Laboratories in the CAP Laboratory Accreditation Program satisfy most of what is necessary to meet the ISO 15189 clause regarding MU. They do so through the following ongoing routines:

- Quality Control (QC)
- Proficiency testing (PT)
- Calibration
- Multi-instrument comparison
- Method comparison
- Generation of data supporting the analytical measurement range as defined by the medical director

CAP accredited laboratories have multiple layers of confirmation that the test is “reliable” and “trustworthy”.

We recommend that laboratories (1) ensure that they have access to the confidence levels for their tests, derived from QC and other analytical processes, and (2) be able to supply a procedure that describes the quality routines that support the validity of the stated confidence levels. See Example of MU Documentation below.

Note: This guideline does not apply to point-of-care testing (POCT).

Detailed CAP Guidance

1. For each of its quantitative tests, a laboratory must have MU values that it can provide if requested.

   Example:

   The MU for Serum Potassium is:

At a level near…	**MU is +/-…**
4.19 mmol/L	0.52 mmol/L; (95% CI = 3.67 - 4.71 mmol/L)
4.69 mmol/L	0.61 mmol/L; (95% CI = 4.08 - 5.3 mmol/L)
7.15 mmol/L	0.84 mmol/L; (95% CI = 6.31 - 7.99 mmol/L)

2. Your laboratory does not have to report out the MU with every result, but it must be able to supply it to any clinician who requests it. The laboratory must define and approve it. The laboratory must set performance requirements for MU, usually expressed as a confidence level within a range.
3. The laboratory should have a procedure that describes the quality routines – such as calibration, QC, PT, multi-instrument comparison, method comparison – that support the validity of the stated confidence levels.
4. The documentation of quality routines and MU calculations should be clear enough that laboratory staff can look it up in the event a clinician calls when the laboratory director/ designee is not present the laboratory.

Sharing information about MU of a test with the laboratory client is an ISO requirement that overarches both clinical consultation and technical consultation, however, the clinical interpretation of a particular result (clinical consultation) must only be conducted by a qualified director/ designee.

There is a critical distinction between two concepts

- *Clinical Consultant* (CAP Laboratory General Checklist – GEN.53650)
- *Technical Consultant Qualifications and Responsibilities* (GEN.53625).

Here is the difference:

- Clinical consultation is primarily a consultation to clinicians regarding the interpretation of a specific test result and its **clinical significance** by a qualified **laboratory director/ designee**
- Technical consultation is primarily a consultation to the laboratory client including analytical details, test methods, performance characteristics, limitations, and quality assurance and does not include interpretation of a particular test for **clinical decision support.**

Here is an example that shows the distinction:

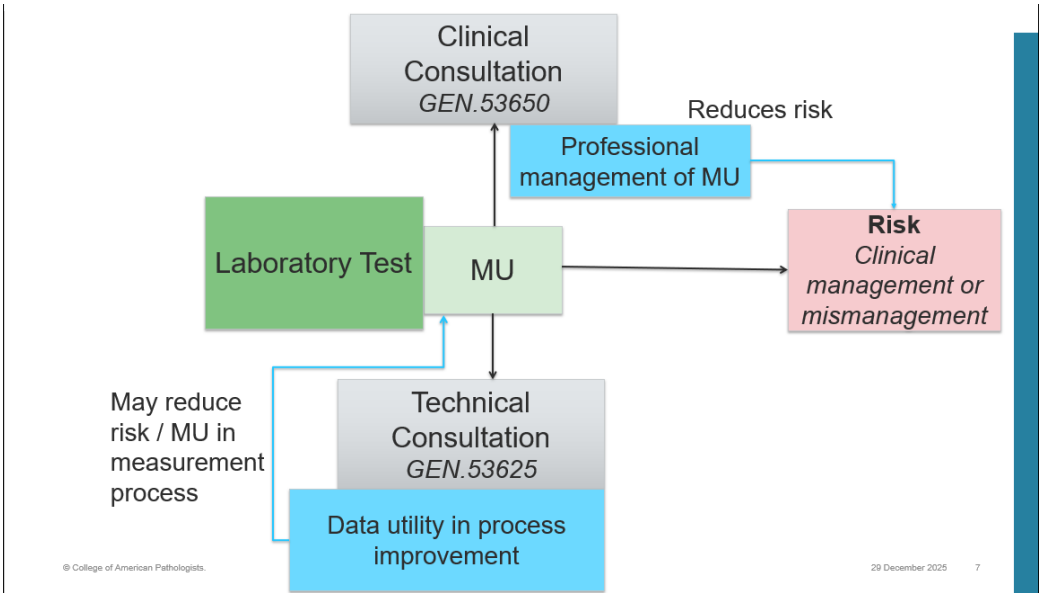
- Technical consultation: explaining significance of MU when values are borderline to a critical limit.
- Clinical consultation: Decide to start treatment or repeat testing based on correlation with clinical manifestations or other guideline-based criteria in conjunction with above borderline laboratory test result.

MU in the above example also addresses the critical association with risk management.

Clinical consultation about MU therefore directly addresses clinical risk of delayed/ missed diagnosis in a case with borderline results.

Potential Risk:

- Clinical consultation- missed diagnosis
- Technical consultation- misinterpret assay limits



MU associated with laboratory test, required consultation, and impacts

5. Clarify what you are including in your MU calculations when you present them to clinicians

Variability caused by (1) pre-examination processes and (2) Biological/ physiologic processes create the largest source of uncertainty in the result from a medical decision standpoint.

Provide an MU assessment most suited to the clinical situation. It is sometimes helpful to cite research on the impact of various variables. AACC/ADLM has published summaries of research.

For cortisol or growth hormones, provide a diagram of the cyclic nature of the levels.

Best Practices

The following are not requirements but are recommended.

1. ***Aim for at least 100 QC measurements when estimating MU***

   The short-term QC CV is inadequate for computing your MU. The laboratory will need sufficient QC data (minimum 20 data points, ideally near 100) to gain representative data on lot changes, instrument maintenance, etc. However, it is not necessary to keep changing the MU for each measurand your laboratory reports.

   Measurand: The specific physical parameter that is intended to be measured (actual truth), unambiguously identified before measurement, distinguishing it from the actual result (estimate).

   Note: The CAP supports the guidance offered by James Westgard: "While there is a 'rule of thumb' that a minimum of 20 controlled measurements should be used to calculate an SD for setting control limits; many more are needed to obtain a reliable estimate of the SD…It would seem prudent to aim for at least 100 measurements when estimating MU." Westgard JO. Basic Quality Management Systems, Madison, WI: Westgard QC, Inc.; 2014, p. 231.

2. ***Re-Calculate MU in response to clinician inquiry.***

   This is a best practice because calibration error (systematic error) and imprecision (random error) vary over time.

   The laboratory may choose to assess MU for a method as part of its method validation/verification workup or may choose to have a process available for calculating a current estimate of MU when required, within a reasonable time frame that allows a timely response to request this information.

   Though not required by the CAP 15189 program, a best practice would be to do both: make an initial estimate of MU during method validation/verification to assess the error that might be seen with the method, and have a process to re-evaluate MU in response to a clinical inquiry.

3. ***Respond to failures in calibration, unacceptable PT results, and QC failures.***

   If calibration fails repeatedly, or you see numerous outliers in your PT, you should reevaluate your MU calculation.

   If you either (1) detect an instrument out of calibration, or (2) fail PT for reasons other than clerical error, this negates the MU assessment, because it calls into question whether you have a stable system.

   If, in the course of testing QC material, after 100 data points, your measurement falls outside the 2 SD range, you should investigate the event. Check with testing personnel for anything unusual, such as QC material that is old, or material that was not mixed well etc. Consider whether the impact was big enough that it might affect MU. With measurements outside the 2 SD range, there is a high probability that something changed with the population or measurement system.

MU may be influenced by routine changes such as changes of reagent batches, different/ new operators, or scheduled instrument maintenance. When you have such changes, you should consider whether you need to recalculate your MU. If you change methodology for the test, you must recalculate your MU as part of the method verification/ validation.

4. ***Get laboratory medical director approval for confidence levels/ analytic measurement range.***

Example of MU Documentation

Here is an example of statistical records and values that would meet the standard for Measurement Uncertainty. Note that there is no requirement to set up your MU calculations in this way.

Measurement Uncertainty Example

Measurement Uncertainty Review Version 1.1 (Month, Year)

1. **Define the measurand: Serum Potassium.**

Assay	Atomic Absorption
Analyzer 1:	Roche-Cobas
Analyzer 2:	
Analyzer 3:	
Analyzer 4:	

Specimen:	Serum
Units:	mmol/L
# decimals:	2
Performance Goal:	Coefficient of variation (or relative uncertainty) is less than 10% (or +/-0.3mmol/L

2. **Calculate the weighted mean and SD for each level QC across all instruments and QC lots.**

LEVEL 1	Analyzer 1		
Month (20XX)	N	Mean	SD
March	26	5.08	0.54
April	6	5.07	0.63
May	17	4.65	0.38
June	18	4.48	0.12
July	18	4.46	0.16
August	20	4.49	0.16

Weighted Stats	N	Mean	SD
Per Analyzer	105	4.69	0.31

LEVEL 2	Analyzer 1		
Month (20XX)	N	Mean	SD
March	26	4.55	0.47
April	6	4.22	0.26
May	18	4.14	0.31
June	18	4.07	0.2
July	19	4.47	0.12
August	20	4	0.14

Weighted Stats	N	Mean	SD
Per Analyzer	107	4.19	0.262

LEVEL 3	Analyzer 1		
Month (20XX)	N	Mean	SD
March	26	7.56	0.72
April	6	7.16	0.47
May	18	7.11	0.76
June	18	6.94	0.25
July	19	6.99	0.16
August	20	6.97	0.13

Weighted Stats	N	Mean	SD
Per Analyzer	107	7.15	0.424

3. **Calculate the combined uncertainty (Uc) of the weighted SD and calibrator if available (or treat as zero)**

4. Determine the expanded uncertainty (U) at a 95%coverage factor (k=1.96)[1]

	Weighted SD	U calibrator	Combined U	k-value	Expanded U
QC Level 1	0.31	0	0.31	1.96	0.61
QC Level 2	0.262	0	0.262	1.96	0.52
QC Level 3	0.424	0	0.424	1.96	0.84

Note 1–The k-value =1.96 is used to create an even 95% confidence deviations actually correspond to 95.4 of all measured results.

5. Determine the relative uncertainty by dividing the combined uncertainty by the weighted mean.

	Weighted Mean	Combined U	Relative U (Coefficient of Variation)	Comment	Meets goal?
QC Level 1	4.69	0.31	6.6%	Method meets acceptable performance if relative CV is less than specified goal	YES
QC Level 2	4.19	0.262	6.3%		YES
QC Level 3	7.15	0.424	5.9%		YES

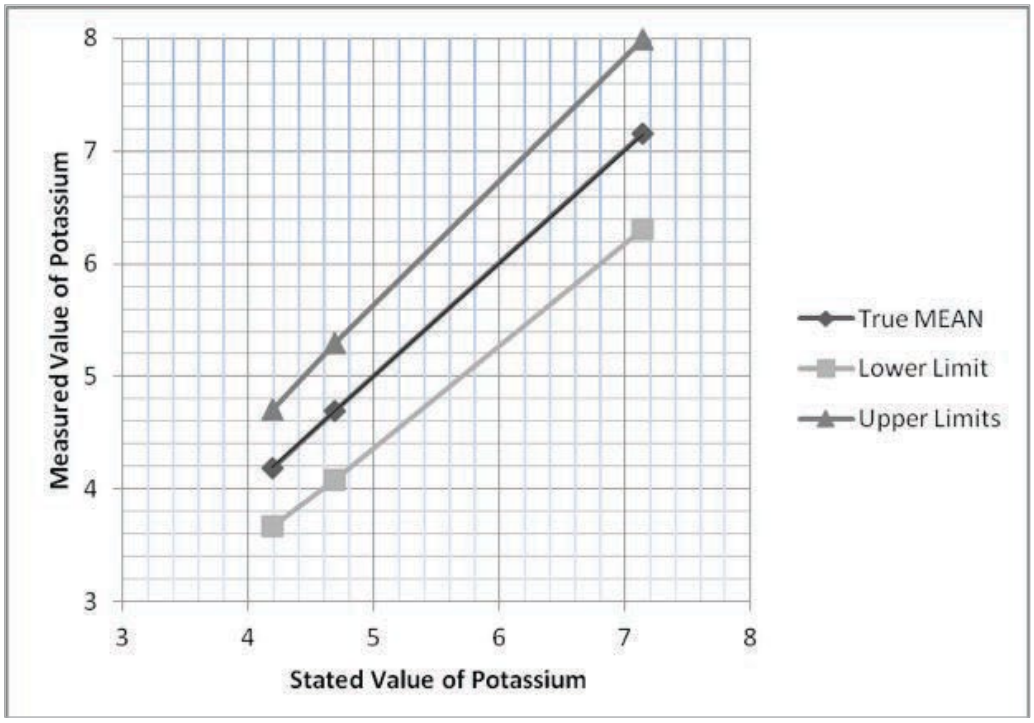
Interpretation of Measurement Uncertainty
The result will be +/- the Expanded Uncertainty with 95% confidence (i.e., correct 19/20 times)

The MU for AA (serum) is:

At a level near 4.19 mmol/L, MU is +/- 0.52 mmol/L; (95% CI = 3.67 - 4.71 mmol/L) At a level near 4.69 mol/L, MU is +/- 0.61 mmol/L; (95% CI = 4.08 - 5.3 mmol/L) At a level near 7.15 mmol/L, MU is +/- 0.84 mmol/L (95% CI = 6.31 - 7.99 mmol/L)

See graph on next page.

True MEAN Values	
4.19	4.19
4.69	4.69
7.15	7.15
Lower Limits of 95% Confidence	
4.19	3.67
4.69	4.08
7.15	6.31
Upper Limits of 95% Confidence	
4.19	4.71
4.69	5.3
7.15	7.99
x	y



	Slope	y-inter
□4.69 Upper	1.18	-0.2342
□4.69 Lower	0.82	0.2342

In simpler terms, we observe a +18% proportional bias at the upper level and a -18% proportional bias at the lower level.

Sample Procedure Supporting MU

Quality Activities Supporting Validity of Serum Potassium Test

The following procedure explains the activities and routines that support the validity of the reported confidence levels for a given test.

Step	Who	What	Detail
1	Phlebotomist / Nurse	Phlebotomy Draw	Competency assessment, at least annually, to prevent hemolysis during blood draw*
2	Testing Personnel	Calibration	Performed every 2 weeks
3	Testing Personnel	QC	Performed daily per shift or every 8 hours
4	All Testing Personnel	PT	Performed 2-3 times per year as required
5	Senior Testing Personnel	Multi-instrument comparison	Performed every 6 months
6	Senior Testing Personnel	Method comparison	Performed every 6 months
7	Supervisor	Training of testing personnel	Provided to testing personnel initially when hired
8	Supervisor	Competency program	Competency verified using the CAP Competency Assessment Program twice during the first year and once annually thereafter
9	Laboratory Medical Director	IQCP for test	Performed once per year or more frequently if warranted by test circumstances or volume

*** Analytical Challenges:** Studies indicate that hemolysis can cause overestimation of potassium results, and specific correction formulas (eg, (K+) – (0.004 x HI) are used to manage this, as discussed in a PubMed study.

Scand J Clin Lab Invest. 2021 Feb;81(1):82-84. doi: 10.1080/00365513.2020.1855363. Epub 2020 Dec 3.

Study of haemolysis interference limit on serum potassium assay on Roche® Cobas 8000 and evaluation of corrected potassium
Claudio Ilardo[1], Amandine Lancien[1], Joël Barthes[1]

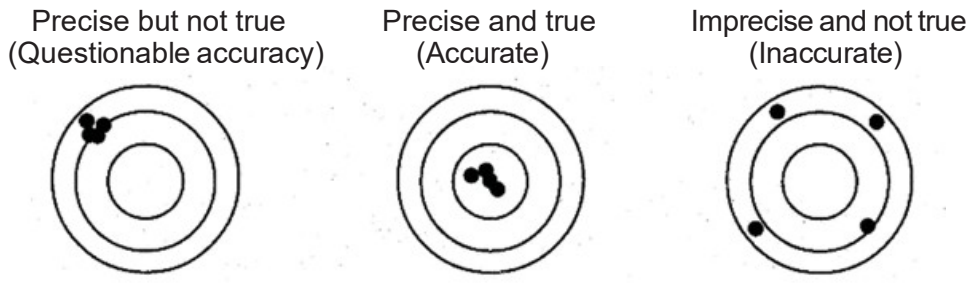
Abstract
Roche Diagnostics® decided a more restrictive haemolysis index (≤20 HI) for approval of the release of serum potassium results. This study examined the risk of overestimating serum potassium results

related to the HI according to the Ricos total error (±5.6%) and evaluated the approach of Martinez-Morillo and Alvarez to determine a corrected potassium. According to Ricos' criteria, our study showed compliant potassium results with HI less than or equal to 75. Between 90 and 100 HI, the results did not show a significant overestimation. The equation to obtain $K_{corrected}$ was: $K_{measured}$ - (0.004 × HI). The use of corrective formulas for adjusting results of potassium could help the laboratory to identify patients at increased risk and to repeat the test as soon as possible.

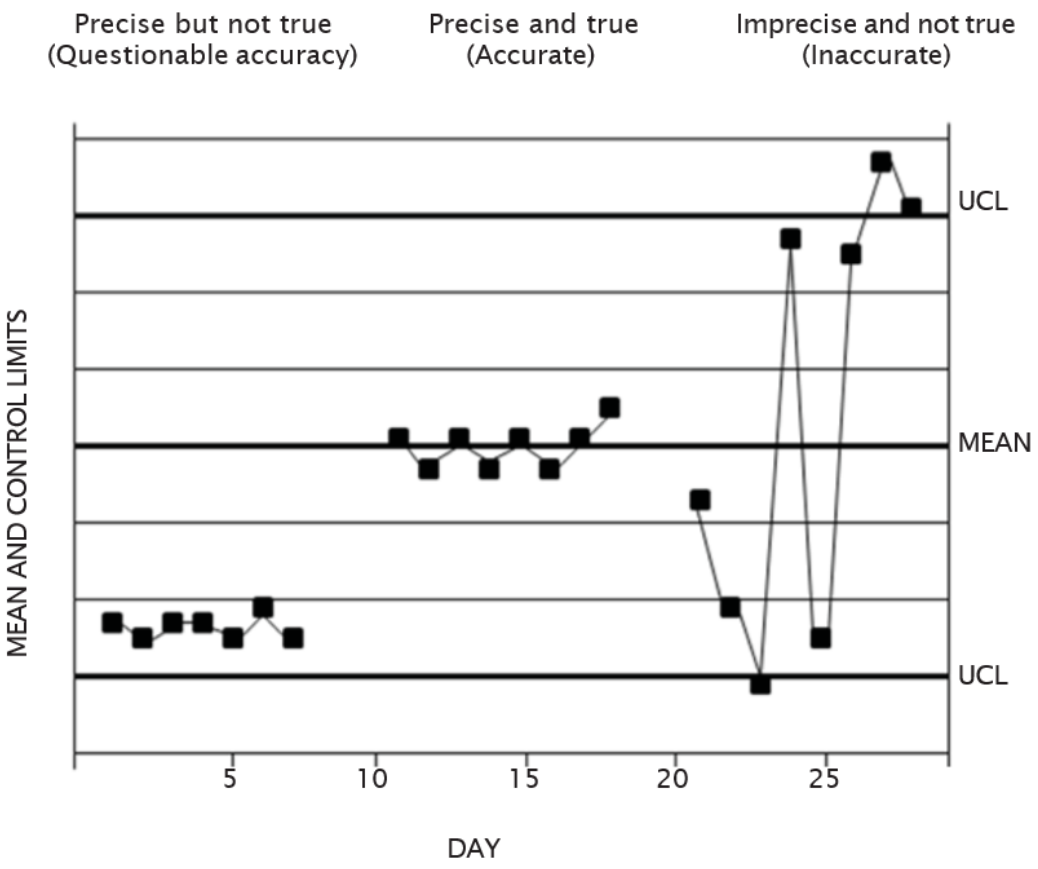
Appendix A – Key Definitions

Term	Definition
Accuracy	The closeness of agreement between a measured quantity value and a true quantity. (VIM 2.13. See Appendix F) It is affected by both trueness and precision of the method.
Precision	Variation between individual measurements performed by replicate testing. This reflects random error of the method and may be estimated by standard deviation or coefficient of variation.
Measurand	The specific physical parameter that is intended to be measured (actual truth), unambiguously identified before measurement, distinguishing it from the actual result (estimate).
Trueness	Closeness of the average of replicate measurand values to its true quantity. This is typically estimated by repeat testing of a sample with an established measure and quantity. It is a reflection of systematic error or bias of the method

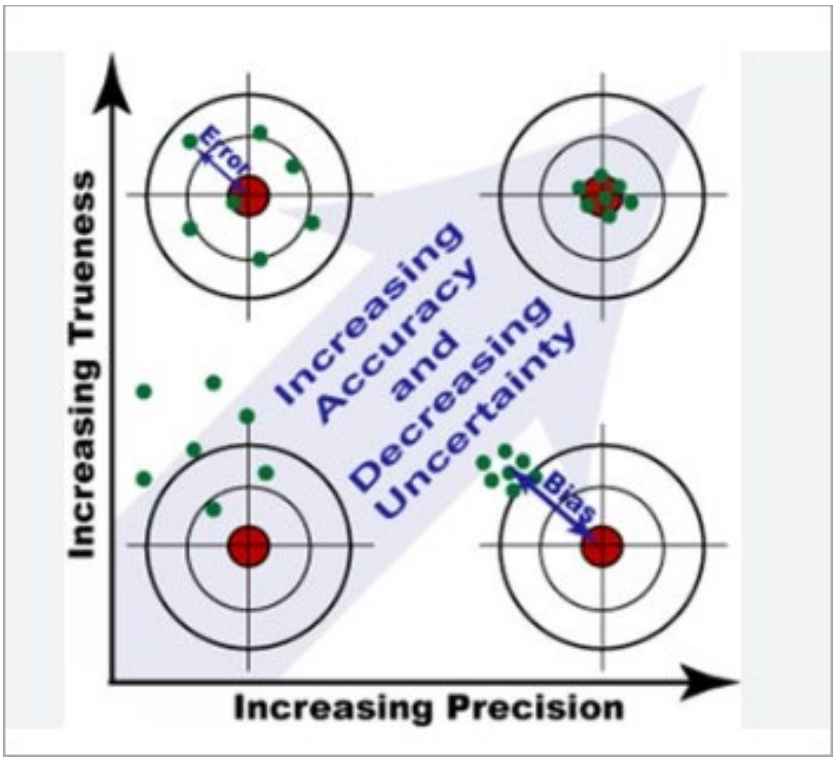
Accuracy versus precision—Target analogy examples



Accuracy vs precision – Control chart examples



Accuracy, precision, trueness, and control chart



Relationship of trueness, precision, accuracy, and measurement uncertainty

Term	Definition
Accuracy	The closeness of agreement between a measured quantity value and a true quantity value of a measurand. (VIM 2.13. See Appendix F.) It is affected by both trueness and precision of the method.
Coefficient of Variation	The ratio of the standard deviation to the mean, expressed as a percentage. Using the CV makes it easier to compare the overall precision of two analytical systems. The CV is a more accurate comparison than the standard deviation as the standard deviation typically increases as the concentration of the analyte increases. Comparing precision for two different methods using only the standard deviation can be misleading. See Appendix D
Confidence Interval	An interval which is expected to typically contain the parameter being estimated. More specifically, given a confidence level, a CI is a random interval which contains the parameter being estimated % of the time.
Measurement Uncertainty	With respect to the measurement system, MU provides a calculable range that provides a context for the clinician to interpret a test result. It is an interval of possible values, within which the analyte is expected to lie, usually with a specified level of confidence. It is a way to assess the reliability of a measurement result by expressing the extent of confidence and range of values associated with it. MU also encompasses factors outside the measurement system, including but not limited to biological variation, that create uncertainty in a test result.
Analytical Measurement Range (AMR)	AMR is the total range of values that a method can measure without modification of a sample, such as dilution or concentration. AMR is usually defined by a linearity experiment. It is the range of values over which there is a linear relationship between measured value and truth (Killeen et al. Arch Pathol Lab Med. See Appendix F).

Term	Definition
Standard Deviation	A statistical measure of dispersion of sample results – based on how much data scatter around the mean.
Variance	Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.

Appendix B – Calculating, Managing, and Expressing MU

MU derives from different sources.

Its calculation and expression also differs by domain – eg, clinical pathology vs. anatomic pathology.

What are the different sources of uncertainty?

MU has traditionally been associated with the instrument or measurement process.

However, in medical laboratory testing we aim to measure analyte as accurately as possible to its true value in vivo (in the patient). This brings additional levels of variability/ uncertainty in measurement that is beyond the scope of the actual testing method/ instrument, usually dependent of the pre-testing conditions.

For medical laboratory testing sources of uncertainty can be classified into:

1. Measurement system factors: instrumental, calibration errors, limitations of measuring devices, methodological, variability in sample preparation, operator skill, statistical, or procedural errors. temperature, humidity, vibrations, or other external conditions affecting measurements during processing.
2. Biological variation: time of sample collection e.g. circadian rhythm affecting cortisol testing, fasting or random glucose testing. degree of hydration (on hematocrit), diet (blood sugar), and drug interactions.
3. Pre-examination phase factors: variations caused in the process of collecting and transporting the sample – eg, tourniquet time for lactic acid, lack of icing for ammonia or blood gas, light sensitivity for bilirubin.

Source 1 of Uncertainty: The Measurement System

Within the domain of the Measurement System, we can break down MU into the following categories:

- Clinical Pathology
- Anatomic Pathology

MU for Clinical Pathology

Measurement uncertainty is the level of uncertainty in the value, when an analyte is measured.

No instrument or test shows absolute perfection – there is always:

- Lack of precision – getting the exact same result on repeat testing, even though minimal
- Lack of accuracy – proximity of the test result to the actual/true value of an analyte on testing

So, when a test is reported, the result carries an inherent uncertainty about the accuracy of the resulted analyte. Understanding MU around a test provides an objective range around the result within which the true value of the analyte exists.

For example:

- Using +/- 2 SD when reporting blood glucose indicates that there is 95% confidence that the true value of blood glucose lies within the provided limits. Or in other words there is 5% chance that the true value of lies outside of the limits.
- The standard for using 2 SD (for 95% confidence interval) or 3 SD (for 99% confidence interval) depends upon the analyte and the clinical situations being managed.

To make good decisions for patient care, a clinician needs to know how accurate a test result is – that is, how far off it may be from the true value.

For example, troponins require higher confidence interval as the results determine a relatively more critical outcome and clinical course in patient management. These standards are usually industry/ national organizations specified or cater to the expectations of the client (clinical or research team).

Measurement uncertainty is not a new concept, and our laboratory practices have inherently been utilizing this information. Labelling the measurement uncertainty at this point helps us better classify the uncertainty about most testing that we performed in medical laboratories. It provides a quantitative indication of the confidence in the result and is a crucial concept in ensuring the quality and reliability of laboratory data.

How do we calculate MU?

MU is a **calculable range** that provides a context for the clinician to interpret a test result.

In other words, it is an interval of possible values, within which the analyte is expected to lie, usually with a specified level of confidence.

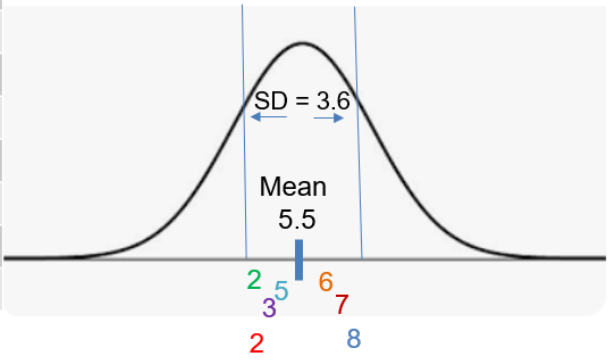
For example:
Question: How certain are we that a glucose result of 262 is really 262?

Answer: We are 95% sure the result is between 258 and 266.

Confidence interval

All MU calculations are based on the concept of **standard deviation (SD)** – a statistical measure of dispersion of sample results – based on how much data scatter around the mean.

Data points	Distance from mean
2	3.5
2	3.5
3	2.5
5	0.5
6	0.5
7	1.5
8	2.5
Mean = 5.5	**SD = 3.6**



Example data collection points, mean, and standard deviation

Note: One might think the SD would be the average distance from the mean, but the actual formula is slightly more complex. Dividing by the number (n) of data points underestimates the dispersion. It is more accurate to divide by n-1. See Appendix C – Formula for standard deviation

Coefficient of Variation (CV) is sometimes preferred by users/ clinicians as a measure for MU over SD.

CV is the ratio of the SD to the mean, expressed as a percentage. In simple words, SD best described absolute uncertainty and CV relative uncertainty. In practice, the value of CV stays relative constant at higher concentration of analyte (reflects stable relative imprecision) and is easier to monitor.

On the other hand, SD stays relatively stable in value at very low analytical measurements when absolute values/ SD are preferred.eg. Lower Limit of Detection/Quantification (**LoD / LoQ)** are frequently expressed as absolute values in chemistry testing, Minimal residual disease monitoring by flow cytometry.

See Appendix D – Comparison of CV and SD.

In the sections below we will walk through some possible methods of calculating MU.

- SD using QC (quality control) testing

- SD plus factor of interlaboratory bias
- SD plus factor of calibration error
- SD plus proficiency testing results

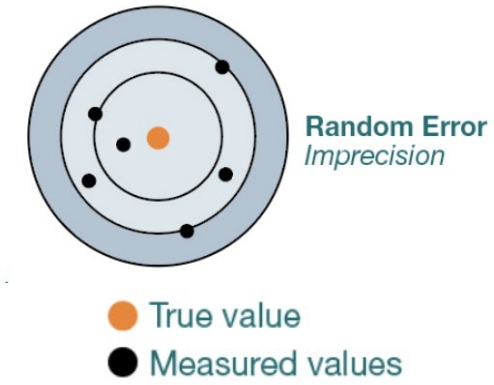
None of these methods yield an exact and true measure of uncertainty, but they all give us ways of providing useful information to a clinician.

SD Using QC

The simplest calculation of MU would be the standard deviation obtained through multiple measurements of QC material.

This captures the **random error** or **imprecision** of the measuring equipment.

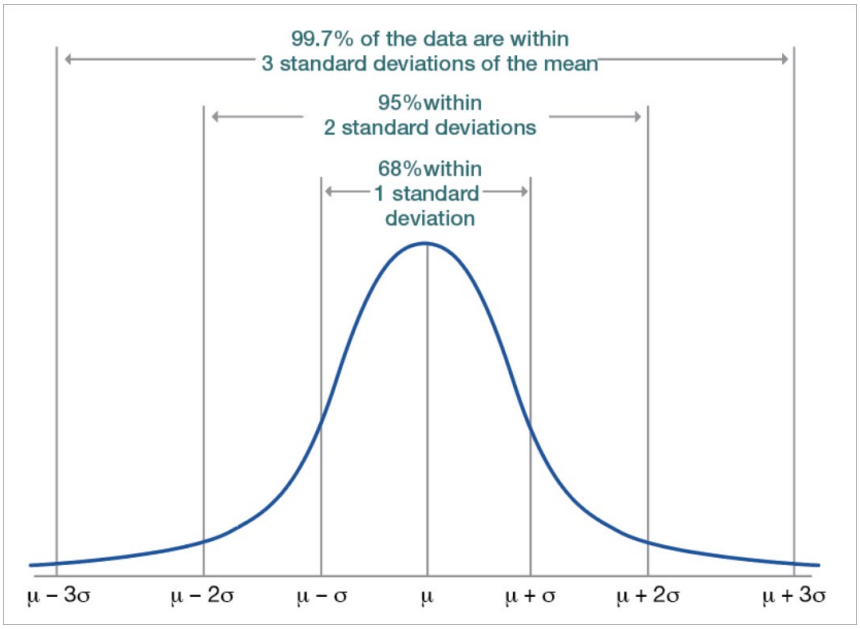
It also takes into account changes in the testing environment and changes in operators.



Random error, instrument imprecision

Based on the assumption that laboratory data generally conform to a normal curve, we can say that the standard deviation gives us 68% confidence that the measurement lies within it.

If we want more confidence, and go for 95% confidence, we multiply the standard deviation by 2 (or 1.96 to be more precise statistically).



Normal (Gaussian) curve – with confidence intervals

These steps can be automated in a spreadsheet or software tool so all we need to do is input the SD. The SD gets multiplied by 1.96 to yield the expanded uncertainty.

	SD	k-value	Expanded Uncertainty
QC Run		1.96	

So, for the glucose example, if the result is 262 and the SD is 2, we can multiply the SD by 2, yielding 4, so the interval is 8.

	SD	k-value	Expanded Uncertainty
QC Run	2	1.96 (roughly 2)	3.96 (roughly 4)

We are 95% sure the result is between 258 and 266.

A glucose measure of 265 a month later would not warrant any change in treatment approach, since the new value lies within the interval.

Since QC material varies, we would calculate confidence intervals for each level or variation of QC material.

Example:

The MU for Serum Potassium is:

At a level near…	**MU is +/-…**
4.19 mmol/L	0.52 mmol/L; (95% CI = 3.67 - 4.71 mmol/L)
4.69 mmol/L	0.61 mmol/L; (95% CI = 4.08 - 5.3 mmol/L)
7.15 mmol/L	0.84 mmol/L; (95% CI = 6.31 - 7.99 mmol/L)

It makes a difference whether we base the calculation on data obtained during the same QC run – or whether it is based on QC runs of the same material on different days.

The latter would have a larger variation. It would also be more realistic in applying to an individual who comes in for a test.

Note that this approach does not take into account:

- Differing testing equipment and methods in different laboratories that measure the analytes
- Different environments in different laboratories
- Variations in the process of collecting and transporting the sample
- Normal biological variation of the person

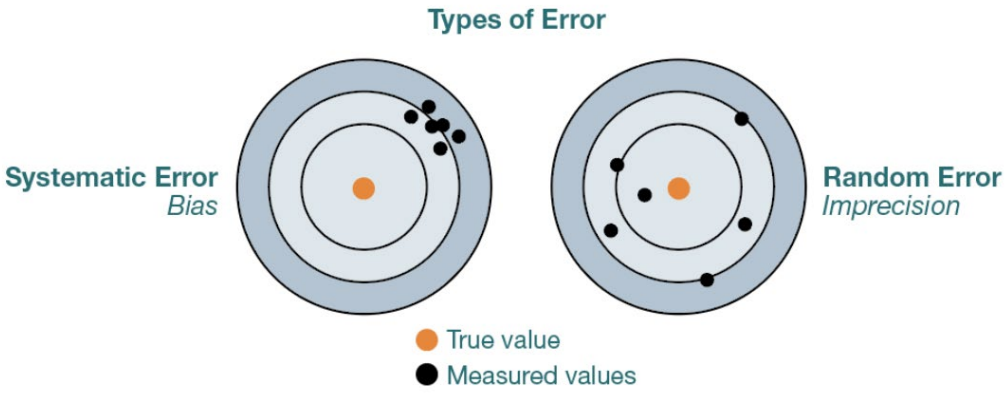
So, we still have additional uncertainty. But focusing on the QC run we have at least provided an objective calculation with a clear logic. The clinician will find this useful in making decisions.

SD plus interlaboratory bias

If we wanted a more robust measure of uncertainty, we could combine:

- The QC statistics of the instrument
- Measures of interlaboratory comparison.

This approach would factor in systematic error, or bias, based on differences in testing location/ environment, equipment, or operators (or all of these factors) as opposed to just random error or imprecision of the measurement instrument.



Systemic error/bias vs. Random error/imprecision

The standard formula – taking into account the SD and the bias – looks like this.

$$U = 2 \times \sqrt{(SD^2 + \left(\frac{Bias}{\sqrt{3}}\right)^2 + SDBias^2)}$$

This formula can be automated in a spreadsheet or software tool so all we need to do is input the SD and the measure of bias. The resulting combined uncertainty multiplied by 1.96 yields the expanded uncertainty.

	SD	Uncertainty from bias	Combined Uncertainty	k-value	Expanded Uncertainty
QC Run	2	.06	2.06	1.96	4.04

Note: The authors wish to recognize the work of Bio-Rad Laboratories, Inc., “Understanding and calculating uncertainty in the clinical laboratory.” 2026. https://www.bio-rad.com/en-us/resources/quality-controls/qc-resources/measurement-uncertainty Graphics, formulas, and content from this report were adapted for this section.

SD and Total Analytical Error (TAE)

When we combine SD and a factor reflecting bias – such as we did in the last section – we get a measure called Total Analytical Error.

This is the error we can measure for the testing or measurement system. It does not factor in variable such as pre-examination issues or biological variation.

Despite its limitations it is a good measure to give to a clinician.

For more information on this approach and the philosophy behind it, see *Laboratory Medicine*, Volume 54, Issue 2, March 2023, Pages 153–159.

SD plus calibration

Another way we can add precision to our calculation of MU is to factor in the strength of the calibration material and protocol we are using for our equipment.

Sometimes manufacturers will publish this and release it with the documents they send out with the calibration material.

If this is available, the calculation would be as follows:

$$U = 2 \times \sqrt{(SD^2) + Cal\ U^2}$$

This formula can be put into a spreadsheet of software tool so all we need to do is input the SD and the measure of bias. This gets multiplied by 1.96 to yield the expanded uncertainty.

	SD	Uncertainty from calibration	Combined Uncertainty	k-value	Expanded Uncertainty
QC Run	2	.05	2.03	1.96	4.02

SD plus Proficiency Testing (PT) / External Quality Assessment (EQA) results

It sometimes makes sense to use the means obtained from PT testing, along with the SD.

PT can be an excellent source to estimate (part of) MU, specifically "bias" (in a "top-down" approach).

1. Martinello F, Snoj N, Skitek M, Jerin A. The top-down approach to measurement uncertainty: which formula should we use in laboratory medicine? Biochem Med (Zagreb). 2020 Jun 15;30(2):020101. doi: 10.11613/BM.2020.020101. Epub 2020 Apr 15. PMID: 32292278; PMCID: PMC7138004.
2. https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf?utm_source=chatgpt.com

Using repeated PT events, you can estimate how far our test results tend to sit from an assigned value (reference/consensus). That becomes a bias uncertainty component.

PT results are typically "single measurements" on different specimen, spaced over time, and influenced by between-lab variation and matrix effects, so they're not an appropriate measure of within-lab imprecision.

Pre-requisites to use PT for MU:
1. PT material is commutable (behaves like a real patient sample) when measured by different laboratory methods. If two methods give the same relationship for patient samples they should give the same relationship for the PT material. If that condition holds, the PT material is clinically relevant.
2. Can be performed if there is the same measurement procedure (method/ platform, calibration traceability, major reagent or software).
3. Sufficient number of PT surveys (~10 to make a meaningful assessment)
4. Be mindful that the "peer consensus" is not necessarily the true value.
5. Determine when to include PT failure/ outlier in the evaluation - ideally when no cause of failure identified/ not corrected on repeat testing on PT survey follow-up.

For example, let's say that on PT, an organization measures vitamin D and there are three antibody assays and one mass spec assay used across 35 laboratories. Let's say the means are as follows:

Method	Mean
Ab-A	23
Ab-B	24
Ab-C	20
mass spec	19

We can say the interval (range) is 19 to 24.

We would add the expanded SD – let's say it's 2 – from the QC of each measurement system (determined separately), so the interval – the measure of uncertainty – is more like 17 to 26.

If a patient notices a jump in vitamin D levels of, say 20 to 24, especially when switching between laboratories the doctor may assure the patient that there is no reason to believe that the true level actually changed; this apparent difference likely represents a change in method with perhaps some additional random variation in the measurement system.

Note: Proficiency testing itself is insufficient to assess the entire MU. It does provide a higher-order (reference) target for its material, which can be used to assess bias/systematic error. This, though, is still at best an indication of the true bias of the method being evaluated.

How can we manage uncertainty for the measurement system in clinical pathology?

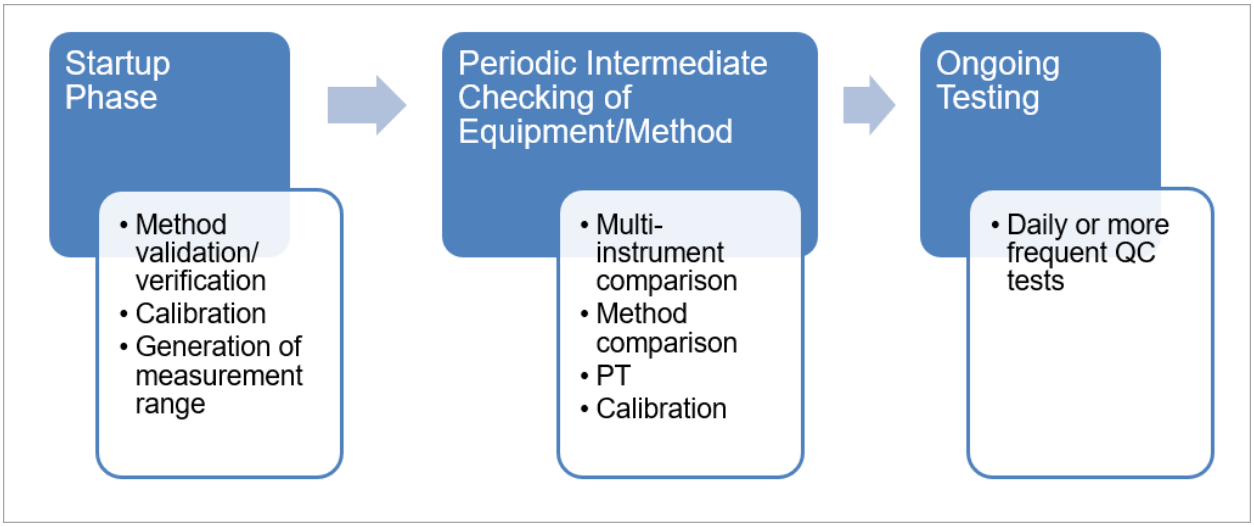
MU cannot be eliminated, but it can be managed and minimized. There is a progression or sequence or processes in a laboratory to minimize and quantify uncertainty.

Here is an overview of the process:

- Startup phase
  - Initial calculation of SD and MU based on initial method validation/verification – including testing of samples for dispersion – to make sure test performs as intended.
  - Generation of data supporting the analytical measurement range as defined by the medical director
    - AMR is the total range of values that a method can measure without modification of a sample, such as dilution or concentration.
    - AMR is usually defined by a linearity experiment. It is the range of values over which there is a linear relationship between measured value and truth (see Killeen et al. Arch Pathol Lab Med.).
    - This is not the same as MU, which is a spot estimate of the error of a result based on known imprecision within the range.

    - Calibration testing
- Periodic Intermediate Checking of Equipment and Testing Method
    - Calibration
    - Multi-instrument comparison
    - Method comparison
    - Interlaboratory/PT testing
- Ongoing Testing phase
    - Daily (or more frequent) QC tests using the equipment

Each of these measurements can serve to confirm that our MU is consistent, or perhaps change the MU.



Processes to manage uncertainty in the measurement system

Sample consultation: MU stemming from the measurement system in clinical pathology

Here is an example:

> **Pathologist**: "Dr. Smith, I wanted to get back to you on your question about the prostate-specific antigen (PSA) test results for your patient. The PSA value came back at 5.2 ng/mL. However, it's important to account for the inherent measurement uncertainty of the assay.
>
> "For this instrument, the coefficient of variation (CV) is about 5% in the range of PSA levels around 5 ng/mL. This means the uncertainty range would be ±0.26 ng/mL (5% of 5.2 ng/mL). Therefore, the true PSA value could fall anywhere between approximately 4.94 ng/mL and 5.46 ng/mL.
>
> "In practice, this uncertainty margin can slightly affect clinical decision-making, especially if the PSA level is near a decision threshold like 4.0 ng/mL. While this particular result is still elevated, we must keep this range in mind when planning further steps, such as recommending a biopsy."
>
> **Clinician**: "Thanks for clarifying. So, while the value suggests further investigation, the uncertainty range reminds me not to overinterpret small variations in monitoring trends."

Note on MU for qualitative tests

PCR testing shows whether a substance is detected, for a given number of cycles.

A laboratory should be prepared to provide the clinician with the data that was used to establish the cutoff values.

For example, the cutoff may come from the instrument information that provides a numerical value, eg, 10 as the cutoff point. Whatever is above the value is considered positive / reactive; whatever is below the value is considered negative / non-reactive.

From the perspective of MU, your uncertainty (2 SD) may be plus or minus 5. Therefore, if the first result is positive, with value of 11, and the clinician asks you to repeat the same sample, and the next result is 9, both fall within the range of plus or minus 5, which is the range of uncertainty.

MU for anatomic pathology

HER-2 stains

Here is an application of MU in semiquantitative HER-2 IHC staining.

The following rules are well established by the CAP/ ASCO guidelines and incorporate the concept of MU at all critical borderline steps or cutoffs.

This also serves as an example of a semiquantitative values with final qualitative results (for which MU must be established for clinically significant step).

In the case of HER-2 testing by IHC, this MU evaluation is already established by the manufacturer for the FDA approved IHC, eg, Roche, like the PATHWAY HER-2 (4B5) IHC.

IHC result of protein expression is used as a surrogate to determine expression levels that would predict gene amplification.

Staining Characteristics	IHC interpretation	MU	Clinical significance
No staining OR Circumferential membrane staining – incomplete, Weak/ faint, and in <10% tumor cells*	0 – Negative	**Low** uncertainty No or – low level IHC expression correlation with non-amplified/non-overexpression Confidence is high that gene is NOT amplified	No benefit from targeted therapy
Circumferential membrane staining – incomplete, weak, and in >10% tumor cells*	+1 Negative	**Low** uncertainty Low IHC expression correlation with non-amplified/non-overexpression	Treatment that targets non-amplified/non-overexpressed levels of HER-2 expression for cytotoxic drug delivery No current recommendation to do FISH testing
Circumferential membrane staining – complete, Moderate to weak, and in >10% tumor cells*	+2 Borderline	Borderline HER-2 expression- the **uncertainty (in predicting amplification) is high** +2 is highest Uncertainty	**Requires** HER-2 FISH to determine gene amplification status
Circumferential membrane staining – complete, Intense, and in >10% tumor cells*	+3 Positive	**Low** uncertainty High IHC expression indicates amplification/ Overexpression Confidence is high that gene IS amplified	Amplified/ overexpressed – Benefit from drugs disrupting HER-2 signaling pathway

Similarly, HER-2 IHC stains may show variability between pathologists.

MU for virology loads

Difficult as it may be to believe, it takes a logarithmic increase in a value for a virology load to signify an actual change.

So, for example, if a viral load measures 100 copies/dL at one instance, and 120 copies/dL the next day, there is no reason to believe there is a clinically significant difference. The MU in these tests is very high and the viral load increases by at least 10 times (~1000 copies/dL) to signify a **true and significant change in viral burden**, rather than measurement uncertainty.

Professional management of MU in microscopic diagnosis

In microscopic evaluation of tissue eg, for lymphoma diagnosis MU depends on

- Factors such as delayed fixation, sampling artifact, and mechanical (hot cautery, crushed tissue) artifact.
- Inter-observer variability (microscopic features being challenging (such as in identification of centroblasts in follicular lymphoma grading)

A pertinent example is that of a scant specimen showing marginal zone lymphoma with rare area rich in larger cells. Here while marginal zone lymphoma is established, a large cell transformation will indicate more aggressive chemotherapy.

A large cell transformation requires sheets/dense cluster of (at least of 5) large cells. In case of a loose cluster of large cells with rare interspersed small cells, or almost good aggregate of large cells in an otherwise scant (needle core) specimen, the diagnostic uncertainty is high.

Further inter-pathologist variability may also be high based on diagnostic skill, confounding activated intermediate sized cells, fixation of specimen, thickness of tissue section, poorly preserved cells due to necrosis.

While MU determination is not quantitative in this case, skilled hematopathologists understand the uncertainty and manage the risk of making the most accurate diagnosis in one of the following ways based on the interpretation of microscopic findings. Which may also show a high degree of interobserver variability.

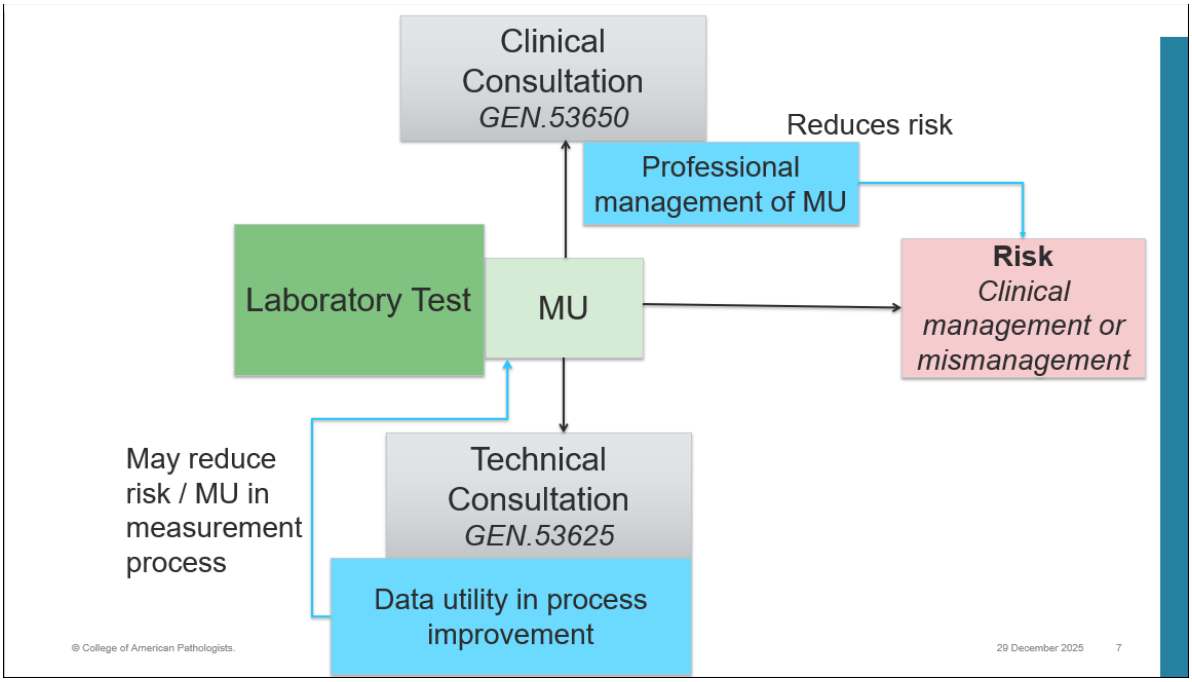
- Focal area rich in large cells does not meet criteria for large cells transformation, however the specimen is scant, and a larger/excisional biopsy may be performed if there is clinical or radiological suspicion of a high-grade process. (MU – low)
- Focal area rich in large cells, concerning for large cells transformation. The specimen is scant and a larger/excisional biopsy is required for definitive evaluation. (MU – high)
- Focal area rich in large cells, shows reasonable clustering, compatible with large cells transformation. (MU – low)

MU (doubt) regarding the large cells transformation in lymphomas is a high-risk situation. Both under- or over-diagnosis is highly harmful in patient care. Thus, professional management of MU, which includes clinical consultation by the pathologist and the review of the patient's laboratory test results, decreases the clinical risk in treatment choices.

Scenario	**Doubt** in diagnosing large cell transformation	MU	Clinical consultation to clinical team by pathologist	Risk management
A	Low (transformation unlikely)	Low	Full excision only if clinically progressive or radiologically (PET avid) concerning	Based on pathology report- risk for underdiagnosis/ and hence undertreatment is low. But if needed, excision may be performed (reducing risk of missing a large cells transformation even further
B	High (cannot rule out transformation)	High	Required to excise the tissue to confirm or exclude large cell transformation	Based on current pathology report- risk for mistreatment is high. Excision is required for confirmation of diagnosis and lower the risk.
C	Low (confirmed transformation)	Low	No need to excise the tissue as large cells transformation is confirmed	Based on diagnosis- risk for underdiagnosis/ undertreatment is low.

Despite different levels of uncertainty in diagnosis, which directly correlates with risk of eventual mis-treatment of patient, the clinical consultation provided by the pathologist manages (modulates) the risk of mistreatment based on additional testing to the lowest.

Diagram showing relation between MU and risk management through clinical consultation (Professional management of MU).



MU associated with laboratory test, required consultation, and impacts

Source 2 of Uncertainty: Biological Variation

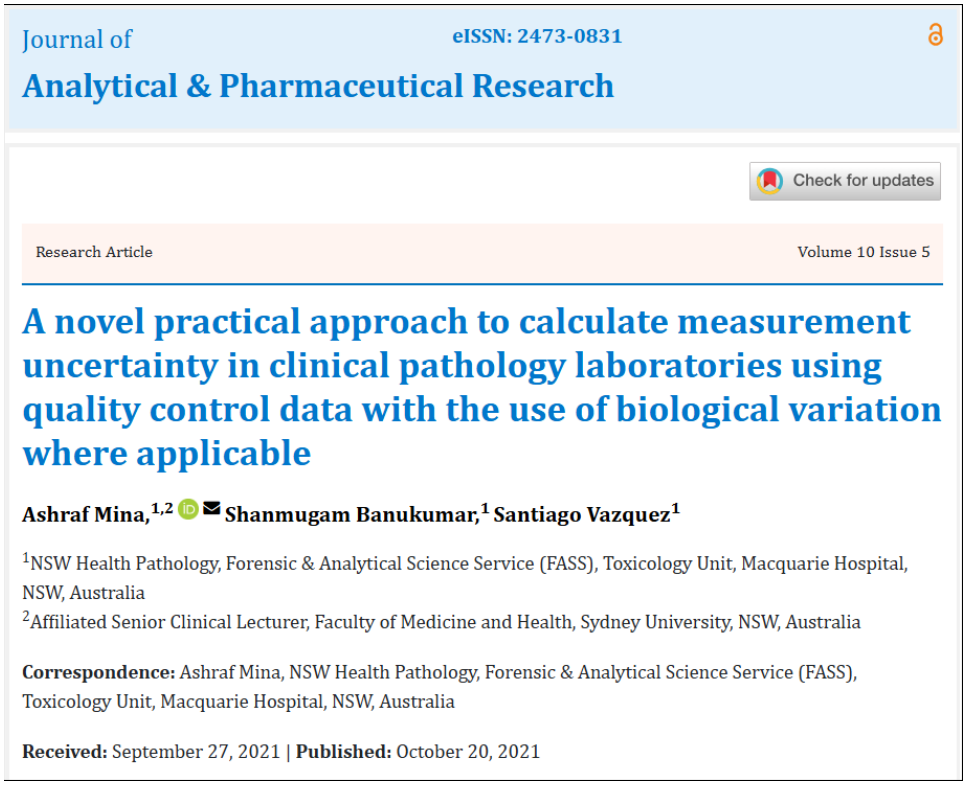
Biological variation (BV) – sometimes called physiological variation – can easily influence a medical test.

Biological variation includes such factors as:

- Diet, hydration
- Fasting
- Cycles
- Drug interactions

Clinicians are interested in estimates of biological variation influencing specific tests, eg, cyclic daily variations in cortisol.

Researchers – for example, Mina and colleagues – have attempted to specify methods that use measures of biological variation, along with MU driven by the measurement system.



Journal of
eISSN: 2473-0831
Analytical & Pharmaceutical Research

Check for updates

Research Article Volume 10 Issue 5

A novel practical approach to calculate measurement uncertainty in clinical pathology laboratories using quality control data with the use of biological variation where applicable

Ashraf Mina,[1,2] Shanmugam Banukumar,[1] Santiago Vazquez[1]

[1]NSW Health Pathology, Forensic & Analytical Science Service (FASS), Toxicology Unit, Macquarie Hospital, NSW, Australia
[2]Affiliated Senior Clinical Lecturer, Faculty of Medicine and Health, Sydney University, NSW, Australia

Correspondence: Ashraf Mina, NSW Health Pathology, Forensic & Analytical Science Service (FASS), Toxicology Unit, Macquarie Hospital, NSW, Australia

Received: September 27, 2021 | **Published:** October 20, 2021

How can we manage uncertainty from biological variation?

Two key methods are:

- Patient instructions
- Ongoing patient monitoring

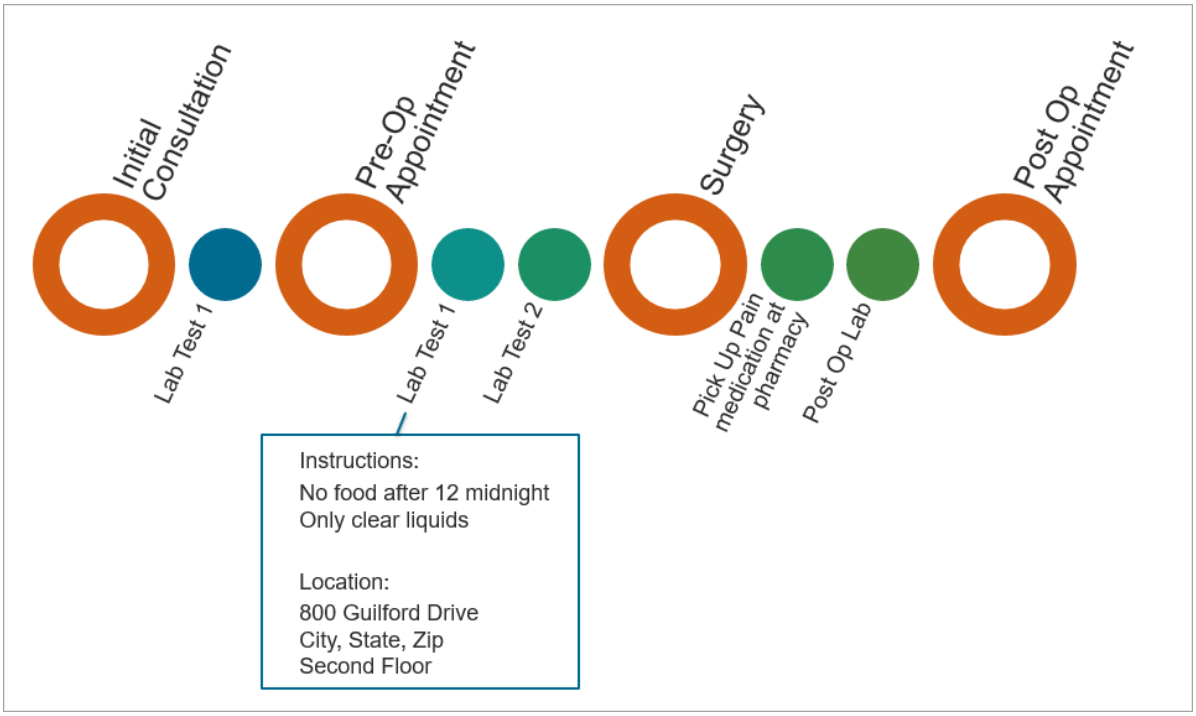
Patient Instructions

It is not uncommon for some patients to not comply with instructions given regarding tests – for example, they may forget to fast before a blood draw, and not tell the laboratory about this. Our testing and interpretation guidelines reflect this fact.

However, it is still important to support compliance with instructions. The more compliant the patient, the more we can reduce uncertainty in the test result.

One way to do this is with clear, complete, and user-friendly instructions – including big picture summaries of when and where the tests will occur, so patients can anticipate them and plan.

Here is an example – a timeline of medical appointments with annotations for testing and preparation details.



Sample of instructions for patient

Instructions such as this improve the overall patient experience with the health system.

Ongoing Patient Monitoring

Ongoing patient monitoring is the practice of repeatedly measuring a patient over time—vitals, labs, waveforms, symptoms, or device data—to understand what's normal for that individual and to detect meaningful change rather than reacting to single data points.

The core problem it addresses is biological variation.

Human biology is noisy—even when nothing is wrong.

Values vary because of:

- Circadian rhythms (day/night effects)
- Hydration, posture, exercise, stress
- Diet and medications
- Normal homeostatic regulation
- Individual "set points" that differ from population averages

So, a single measurement can be misleading:

- A value may be abnormal but harmless for that person
- Or within reference range but dangerous for them

Monitoring helps us separate signal from noise.

Here is how ongoing patient monitoring relates to MU:

1. MU defines how much "wiggle" a result can have.

Measurement uncertainty describes the expected range around a test result within which the true value is believed to lie.

In practice:

- Every laboratory result = measured value ± uncertainty
- Small changes may fall inside MU → likely noise
- Larger or consistent changes may exceed MU → likely real biological change

MU sets the boundary between noise and signal.

2. Ongoing monitoring turns MU from abstract math into clinical meaning

A single result:

- Cannot be confidently interpreted if the change is within MU

Repeated results:

- Allow clinicians to see whether changes are:
    - Random (within MU, non-directional)
    - Systematic (directional, persistent, exceeding MU)

Monitoring answers the key question MU raises:

"Is this difference real, or could it be measurement variability?"

3. Reference Change Value (RCV) is where MU and monitoring meet.

RCV combines:

- Analytical variation (MU / CV_a)
- Biological within-person variation (CV_i)

RCV answers:

"How much must a result change between two measurements before we can be confident the change is real?"

Ongoing monitoring:
- Provides the serial data needed to apply RCV meaningfully
- Prevents overreaction to changes smaller than expected MU + biologic variation

For an online calculator of RCV, here is a site: EFLM Biological Variation

4. Monitoring reduces the risk of misinterpretation caused by MU

Without monitoring:
- A single borderline result may trigger:
  - Unnecessary intervention
  - Repeat testing
  - Anxiety or alarm

With monitoring:
- Results can be interpreted in context:
  - Is the value stable within expected MU?
  - Is it drifting beyond MU over time?
  - Is the rate of change clinically relevant?

Monitoring is a risk control for MU-related misinterpretation

5. MU explains why "normal range" is not enough

Reference intervals:
- Are population-based
- Ignore individual baselines
- Do not account for MU in isolation

Ongoing monitoring:
- Allows patient-specific thresholds
- Uses MU-aware interpretation rather than hard cutoffs

Example:
- Two results within reference range
- But a change exceeding MU/RCV
- → Clinically significant despite being "normal"

6. ISO 15189 framing (plain English intent)

From an ISO perspective:
- MU characterizes analytical reliability
- Monitoring manages residual uncertainty
- Together, they support:
  - Valid clinical interpretation
  - Risk-based decision-making
  - Patient safety

MU alone does not ensure good decisions.
Monitoring alone ignores analytical limits.
They must work together.

Sample consultation: MU stemming from physiologic factors

Pathologist: "In response to your question, the PSA level for your patient with benign prostatic hyperplasia came back at 6.8 ng/mL. However, before interpreting this as an indication for possible malignancy, we should consider biological factors that could affect PSA levels.

"For example, the patient recently underwent a digital rectal exam (DRE) and has reported mild prostatitis. Both of these conditions may transiently elevate PSA levels – there is one study from 2003 from European Urology Journal that indicates prostatitis can elevate PSA levels from 30 to 100%. For this patient, even if we take the low end of 30%, this could mean an elevation of about 1.56 ng/mL, putting the baseline PSA in between 3.64 to 6.76 ng/mL

"This introduces another layer of uncertainty beyond the analytical measurement error. It might be prudent to repeat the PSA test after the inflammation resolves or ensure no further exacerbating factors like recent ejaculation, which can also increase PSA levels."

Clinician: "That's a good point. I'll reassess the patient's history and current symptoms and consider repeating the PSA in a few weeks to get a clearer picture."

Pathologist: "Exactly. Balancing measurement uncertainty with biological variability can help us make better-informed decisions without jumping to unnecessary interventions."

Source 3 of Uncertainty: Pre-examination phase factors

While harder to quantify than measurement system factors, it is widely acknowledged that pre-examination factors drive considerable uncertainty and are the single biggest source of errors in medical tests.

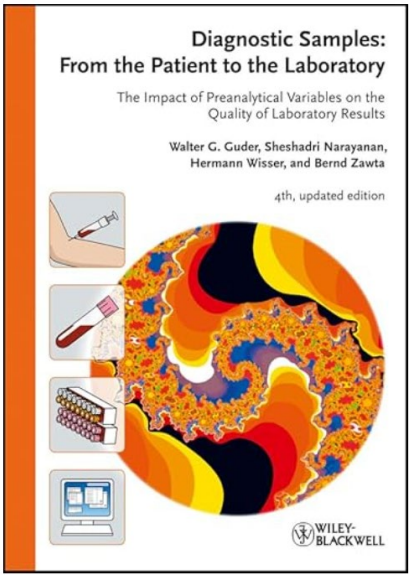
Pre-examination processes include:
- Specimen collection
- Specimen transport
- Specimen receiving

For example, specimen collection using torniquet may result in high lactate ammonia.

Blood draws from IV lines from hospital patients can cause variability in test results, if not properly collected.

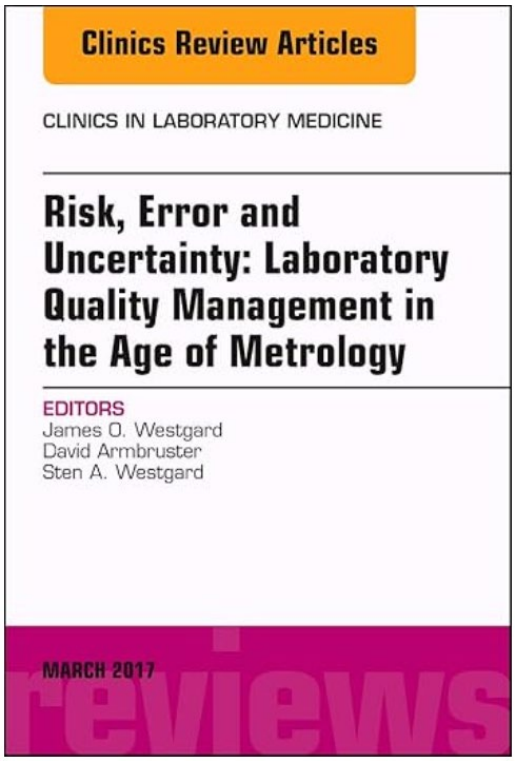
Many references are available that explain the mechanism and effect of pre-examination procedures on laboratory tests. Here are some book suggestions:

1. **Diagnostic Samples: From the Patient to the Laboratory: The Impact of Preanalytical Variables on the Quality of Laboratory Results, 4th, Updated Edition**



Walter G. Guder, Sheshadri Narayanan, Hermann Wisser, Bernd Zawta
https://www.wiley.com/en-ie/Diagnostic+Samples%3A+From+the+Patient+to+the+Laboratory%3A+The+Impact+of+Preanalytical+Variables+on+the+Quality+of+Laboratory+Results%2C+4th%2C+Updated+Edition-p-9783527691081

2. **Risk, Error and Uncertainty: Laboratory Quality Management in the Age of Metrology, An Issue of the Clinics in Laboratory Medicine, 1st Edition**



By James O. Westgard, Ph.D., FACB, David Armbruster and Sten Westgard, MS

https://www.asia.elsevierhealth.com/risk-error-and-uncertainty-laboratory-quality-management-in-the-age-of-metrology-an-issue-of-the-clinics-in-laboratory-medicine-9780323477437.html?utm_source=chatgpt.com

How can we manage uncertainty from pre-examination phase factors?

Two key methods are:

- Internal auditing of processes
- Monitoring of metrics

Internal audits can help by

- Discovering ways in which the pre-examination processes are not operating as designed – this creates a risk of inaccurate test results
    - Determine if temperatures were stabilized during transport
    - Determining whether laboratory staff took temperatures as required and acted on any outliers
- Identifying risks in the process
- Identifying opportunities for improvement

Monitoring of metrics is helpful because metrics can provide “early warning” of future problems, eg.

- Training/competency assessment data can show a trend toward inadequate training of staff – this creates a risk of improper handling of specimens and inaccurate test results
- Trends in QC data can show drifting of results in a certain direction – which may be caused by not following pre-examination protocols

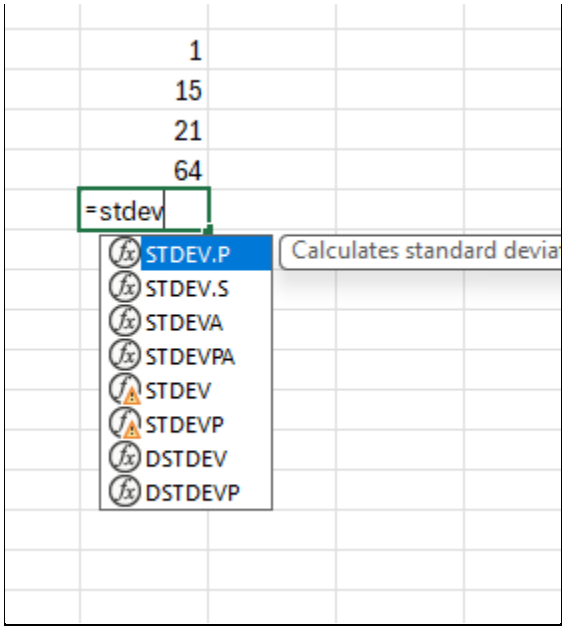
Sample consultation: MU stemming from pre-examination phase factors

Here is a sample consultation:

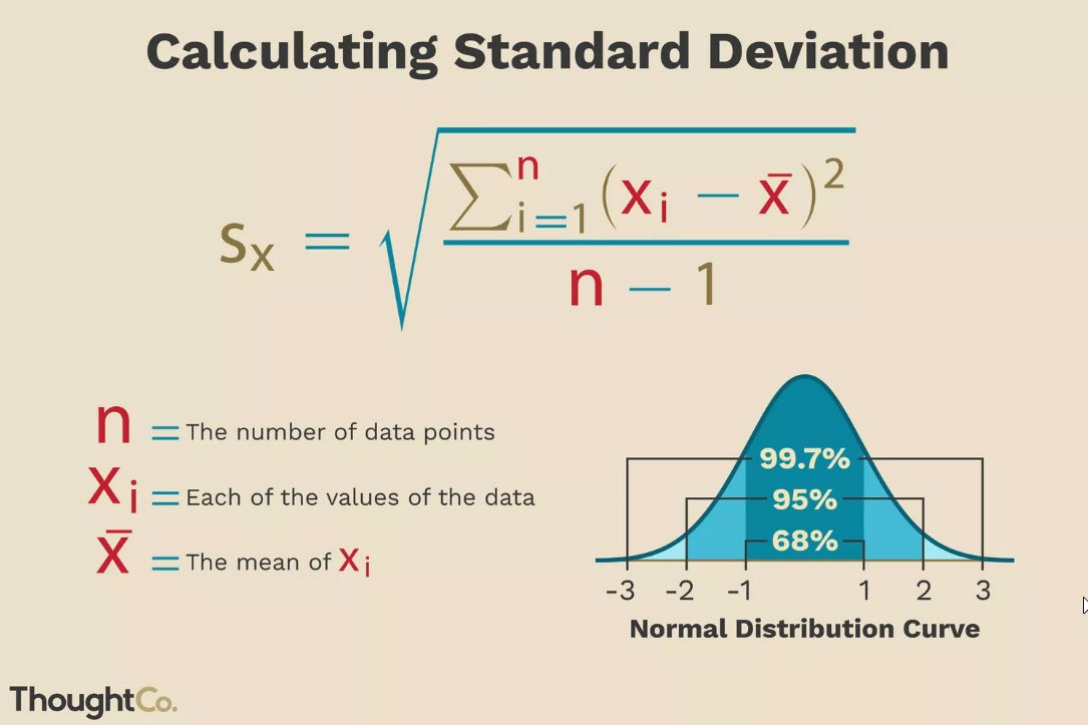
> **Pathologist**: “Dr. Smith, I wanted to answer your question about the lactate and ammonia results for our patient. The lactate returned at **3.2 mmol/L**, and the ammonia at **78 µmol/L**. While both values are above the reference range, it’s important to consider **pre-examination sources of measurement uncertainty**, particularly those related to specimen collection and handling.”
>
> “Both lactate and ammonia are highly sensitive to **tourniquet time, fist clenching, and delays in sample processing**. Prolonged tourniquet application or local ischemia can increase anaerobic metabolism in the sampled limb, leading to **artifactual elevation**. For lactate, short periods of venous stasis can increase measured levels by **10–30%**. Similarly, delayed transport or inadequate cooling allows ongoing cellular metabolism, further increasing measured concentrations.”
>
> “Under optimal conditions, lactate and ammonia results are reasonably stable. However, when pre-analytical conditions are suboptimal, the observed value may not reflect the patient’s true systemic level. and ammonia can rise even more rapidly if the sample is not immediately placed on ice and processed promptly.”
> “In this case, the phlebotomy notes indicate a **prolonged tourniquet time** and a **delay to centrifugation**, which introduces an uncertainty that tends to falsely elevate the result.”
>
> **Clinician**: “That’s helpful. I’ll interpret these results cautiously and repeat both tests with minimized tourniquet time and rapid transport on ice, rather than escalating care based on a single potentially pre-analytical artifact.”

Appendix C – Formula for SD

Here is the Excel formula for calculating standard deviation.



Excel automates the mathematical formula.



In this standard deviation formula, "n-1" is used instead of "n" because when calculating the standard deviation from a sample, dividing by "n-1" provides a more accurate estimate of the population standard deviation by

correcting for the bias introduced when using the sample mean to estimate the population mean; this is known as "Bessel's correction."

Key points about using "n-1":

- Sample vs. Population:
  - When you only have a sample of data from a larger population, the sample mean will likely be closer to the data points in your sample than the true population mean, leading to an underestimation of the actual variance if you divide by "n".
- Degrees of Freedom:
  - By subtracting 1 from "n", you are essentially accounting for the "loss of information" due to the estimation of the population mean from the sample mean.

Appendix D – Comparison of CV and SD

The coefficient of variation is the ratio of the standard deviation to the mean. Express CV as a percentage. Use the following formula to calculate the CV:

$$CV\ \% = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

Using the CV makes it easier to compare the overall precision of two analytical systems. The CV is a more accurate comparison than the standard deviation as the standard deviation typically increases as the concentration of the analyte increases. Comparing precision for two different methods using only the standard deviation can be misleading.

Example Scenario:

Compare a hexokinase method and glucose oxidase method for measuring glucose. The standard deviation for the hexokinase method is 4.8. The standard deviation for the glucose oxidase method is 4.0. Based on the standard deviation, you might conclude the glucose oxidase method is more precise than the hexokinase method.

However, in this example, a comparison of the CV shows the methods are equally precise. Assuming the mean for the hexokinase method is 120 and the mean for the glucose oxidase method is 100, the CV for both methods is 4%.

Example of Coefficient of Variation (CV) versus Standard Deviation (SD)

Analyte: Glucose
Method: Hexokinase
Standard deviation = 4.8
Mean = 120

Analyte: Glucose
Method: Glucose oxidase
Standard deviation = 4.0
Mean = 100

Which method is more precise?

Calculate the CV to see:

$$CV\ (\%) = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100$$

$$\frac{4.8\ (SD)}{120\ (Mean)} \times 100 = 0.04\%\ (CV) \qquad \frac{4.0\ (SD)}{100\ (Mean)} \times 100 = 0.04\%\ (CV)$$

In this example, the CV proves both methods are equally precise.

Appendix E – Comparison of MU Requirement in ISO 15189:2012 and ISO 15189:2022

Requirements:

ISO 15189:**2012**	New Requirements – ISO 15189:**2022**
Maintain MU values	same
Compare MU against performance specifications	same
Review MU evaluations regularly	same
	Document the rationale for exclusions when MU is not possible or relevant
Make MU available to laboratory users upon request	same
	In answering inquiries, take into account sources of uncertainty beyond measurement uncertainty, such as but not limited to biological variation
	Estimate MU for qualitative tests based on defined cutoff values
Consider MU in non-quantitative testing. Evaluate MU in the significant intermediate steps of testing.	For qualitative examinations, include MU from intermediate quantitative steps or IQC data in high-risk parts of the process. In 2022, the guideline specifies more clearly the need to monitor MU in the intermediate quantitative step of high significance. These two are partly overlapping.
	When relevant, consider MU during method verification or validation.

5.5.1.4 Measurement uncertainty of measured quantity values

The laboratory shall determine measurement uncertainty for each measurement procedure in the examination phase used to report measured quantity values on patients' samples. The laboratory shall define the performance requirements for the measurement uncertainty of each measurement procedure and regularly review estimates of measurement uncertainty.

NOTE 1 The relevant uncertainty components are those associated with the actual measurement process, commencing with the presentation of the sample to the measurement procedure and ending with the output of the measured value.

NOTE 2 Measurement uncertainties may be calculated using quantity values obtained by the measurement of quality control materials under intermediate precision conditions that include as many routine changes as reasonably possible in the standard operation of a measurement procedure, e.g. changes of reagent and calibrator batches, different operators, scheduled instrument maintenance.

NOTE 3 Examples of the practical utility of measurement uncertainty estimates might include confirmation that patients' values meet quality goals set by the laboratory and meaningful comparison of a patient value with a previous value of the same type or with a clinical decision value.

The laboratory shall consider measurement uncertainty when interpreting measured quantity values. Upon request, the laboratory shall make its estimates of measurement uncertainty available to laboratory users.

Where examinations include a measurement step but do not report a measured quantity value, the laboratory should calculate the uncertainty of the measurement step where it has utility in assessing the reliability of the examination procedure or has influence on the reported result.

Language from ISO 15189:2012

7.3.5 Evaluation of measurement uncertainty (MU)

i) The MU of measured quantity values shall be evaluated and maintained for its intended use, where relevant. The MU shall be compared against performance specifications and documented.
NOTE ISO/TS 20914 provides details on these activities together with examples.

j) MU evaluations shall be regularly reviewed.

k) For examination procedures where evaluation of MU is not possible or relevant, the rationale for exclusion from MU estimation shall be documented.

l) MU information shall be made available to laboratory users on request.

m) When users have inquiries on MU, the laboratory's response shall take into account other sources of uncertainty, such as, but not limited to biological variation.

n) If the qualitative result of an examination relies on a test which produces quantitative output data and is specified as positive or negative, based on a threshold, MU in the output quantity shall be estimated using representative positive and negative samples.

o) For examinations with qualitative results, MU in intermediate measurement steps or IQC results which produce quantitative data should also be considered for key (high risk) parts of the process.

p) MU should be taken into consideration when performing verification or validation of a method, when relevant.

Language from ISO 15189:2022

Appendix F – References and Resources for Further Study

1. Joint Committee for Guides in Metrology. JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition 2012 version with minor corrections. JCGM website. Available at: http://www.bipm.org/en/publications/guides/ Accessed January 8, 2026.

2. Westgard JO. *Basic Quality Management Systems*, Madison, WI: Westgard QC, Inc.; 2014.

3. Anthony A. Killeen, Tom Long, Rhona Souers, Patricia Styer, Christina B. Ventura, and George G. Klee (*2014*) Verifying Performance Characteristics of Quantitative Analytical Systems: Calibration Verification, Linearity, and Analytical Measurement Range. Archives of Pathology & Laboratory Medicine: September 2014, Vol. 138, No. 9, pp. 1173-1181. Available at: http://dx.doi.org/10.5858/arpa.2013-0051-CP Accessed January 8, 2026.

4. ISO/TS 20914:2019 *Medical laboratories — Practical guidance for the estimation of measurement uncertainty.*
   Published (Edition 1, 2019) This publication was last reviewed and confirmed in 2023.
   https://www.iso.org/standard/69445.html Accessed January 8, 2026.

5. Martinello F, Snoj N, Skitek M, Jerin A. The top-down approach to measurement uncertainty: which formula should we use in laboratory medicine? Biochem Med (Zagreb). 2020 Jun 15;30(2):020101. doi: 10.11613/BM.2020.020101. Epub 2020 Apr 15. PMID: 32292278; PMCID: PMC7138004.

6. Selection, Use and Interpretation of Proficiency Testing (PT) Schemes. Eurachem. Second Edition. 2011.
   https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_PT_Guide_2011.pdf?utm_source=chatgpt.com Accessed January 8, 2026.

7. Bio-Rad Laboratories, Inc., “Understanding and calculating uncertainty in the clinical laboratory.” 2026.
   https://www.bio-rad.com/en-us/resources/quality-controls/qc-resources/measurement-uncertainty
   Accessed February 26, 2026.

Appendix G – Summary Poster

This summary poster is available through the College of American Pathologists. Contact: QMEd@cap.org

COLLEGE of AMERICAN PATHOLOGISTS
Laboratory Quality Solutions

Measurement Uncertainty: It's More Than a Number

It encompasses numerical and qualitative uncertainty, including quality routines that give us confidence in our testing.

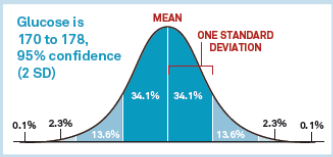
What *creates* uncertainty?

Measurement system factors

- Instrument fluctuations
- Environment factors, such as temperature, humidity, vibration
- Different operators and methods
- Calibration tools

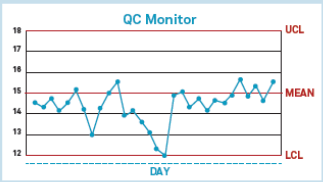
How can we *express* uncertainty?

- Calculable range based on the standard deviation and confidence interval
- Description of all the measures taken to assess, optimize, and manage the test system, plus any results
- Different ways of demonstrating that a test is reliable

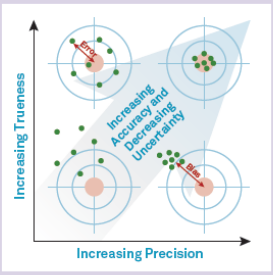


How can we *manage* uncertainty?

- PT, QC
- Instrument comparison
- Method comparison
- Calibration
- Training



Key MU Concepts



Other Influences on Results and Interpretation

Biological variation

- Diet, hydration
- Fasting
- Cycles
- Drug interactions

Estimates of biological variation influencing specific tests
(eg, cyclic daily cortisol variations)

- Patient instructions
- Ongoing patient monitoring

Pre-examination phase factors

- Specimen collection
- Specimen transport
- Specimen receiving

Estimates of effects of procedures
(eg, specimen collection using a tourniquet may result in elevated lactate or ammonia)

- Internal process audits
- Metric monitoring
- Training

Term	Definition
Measurement Uncertainty	The level of uncertainty in the value when an analyte is measured
Accuracy	The closeness of agreement between a measured quantity value and a true quantity
Precision	Variation between individual measurements performed by replicate testing
Trueness	Closeness of the average of replicate measurand values and its true quantity Typically estimated by repeat testing with an established measurand quantity

For more information about MU, see the CAP 15189 Measurement Uncertainty Guide. Go to cap.org > Accreditation > CAP 15189