Dr. Graham White provides an explanation of measurement uncertainty (MU), its role in the clinical laboratory, and the advantages it has over the use of total error.

The Hitch-hiker's Guide to Measurement Uncertainty (MU) in Clinical Laboratories

April 2012

Graham White
SA Pathology, Flinders Medical Centre, Bedford Park, Adelaide, SA 5042 Australia
Graham.White@health.sa.gov.au

- What types of laboratory measurement should use MU?
- Why does MU matter in routine clinical laboratories?
- Can the Total Error (TE) approach help?
- Limitations of the Total Error Concept (TE)
- What is Measurement Uncertainty (MU)?
- What does MU include?
- How does MU handle bias?
- How is MU estimated?
- What if Bias is considered significant?
- How is MU calculated for a measurement calculated from several other results?
- How should MU estimates be assessed?
- Summary
- Should MU be routinely reported to clinicians?
- Summary of MU and TE
- References

I appreciate the invitation to provide further discussion on measurement uncertainty to follow the recent article by James Westgard on the recent CLSI MU guideline. A hitch-hiker likes to keep things simple, have an easily understood map and reach the destination with minimal discomfort. This hitch-hiker's destination is to briefly demonstrate to routine laboratorians, using minimal jargon and statistics, that MU is logical, easy to understand, useful and appropriate to implement in clinical laboratories [1,2].

What types of laboratory measurement should use MU?

All types of measurement that have a magnitude expressed as a number (e.g. 4.6, 1×10^{-9}) and a reference (e.g. mmol/L, dimension one) e.g. measurement of plasma calcium concentration, white blood cell count, number of CAG nucleotide repeats.

Why does MU matter in routine clinical laboratories?

Clinicians compare most measurement results with reference values and with previous results from the same patient. Results should therefore be reliable and accurate, but in practice they suffer from error. When verifying the performance characteristics of a routine measurement procedure, repeatability experiments are usually performed i.e. replicate measurements of the same sample with conditions kept as constant as possible. If the measuring system is sufficiently sensitive, a range of different results will usually be obtained. Which is the true result for the sample? We obviously can't say, but clearly the results must contain some error, and the magnitude of error is not the same for the differing results. There is therefore uncertainty as to what the true value is. A dispersion of results is similarly obtained if a patient sample is repeatedly measured under replicate conditions.

A patient report shows serum rhubarb results on two samples collected a week apart as 3.1 mmol/L and 3.3 mmol/L; upper reference value: 3.0 mmol/L. The clinician asks the laboratory: Is the first result definitely high and is the second result really higher than the first? The laboratory can't answer without having some quantitative knowledge about the measurement uncertainty associated with each of the results.

Can the Total Error (TE) approach help?

The TE approach identifies systematic error (Bias) and random error (Imprecision) as the two components of total measurement error. Bias is a predictable offset of results from a reference value, usually estimated as the difference between a reference value and the mean result obtained when the reference is measured in replicate by the routine measurement procedure. The reference value chosen may be a certified reference material, a peer group mean in an EQA etc. Bias therefore has a value which can be used to eliminate or minimise the offset e.g. by re-calibration or by adjusting raw results with a correction factor.

The magnitude of imprecision is unpredictable for each measurement result produced by an assay, due to factors such as fluctuations in electro-mechanical performance, reagent and calibrator batch changes, different operators, routine instrument maintenance. Imprecision is usually estimated by measuring QC materials in different analytical runs spread over sufficient time to include as many as possible of the above routine changes in measurement conditions. The dispersion of results obtained over time for the same QC batch is considered to approximate to a Gaussian (normal) distribution, so that the magnitude of the dispersion can be statistically quantified as a standard

deviation (SD) from the mean value. Laboratories make the assumption, which is not always true, that patient samples behave as do QC samples in a measuring system, so that the dispersion of results is expected to be similar if a patient's sample was to be repeatedly measured i.e. have similar imprecision.

The Total Error Concept describes total error of a measuring system as: TE = Bias + 1.65 SD, where 1.65 SD represents the ~95 % dispersion of results obtained on one side of the imprecision Gaussian curve.

Limitations of the Total Error Concept (TE)

There are several problems with the Total Error approach when applied to individual results:

1. A bias value cannot be exactly known. For example, analyte values assigned to certified reference materials are inexact, being expressed in the form of x ± y units. In addition, a mean value obtained for the reference material from replicate measurements by the routine assay has unavoidable imprecision, so that the mean value using the routine procedure is also in the form of x ± y units. As both values used for calculating the bias have uncertainty, the bias value cannot be exactly known. Similarly, the random error is just that, so it cannot be exactly known for an individual measurement result. Therefore, the total error of a measurement result cannot be exactly known. This is unavoidable even with state-of-the-art reference measurement procedures, which is why analyte values assigned to certified reference materials is stated in the form of x ± y units.

2. The calculation of total error is Bias value + 1.65 SD. If a value for the bias in a measurement procedure is known, why does the Total Error approach include it in the calculation of total error rather than eliminate or minimise it by either re-calibration or application of a correction factor to raw results? If a bias value is unobtainable, then bias is unknown and cannot be addressed.

3. The TE calculation adds a bias value to a probability distribution (1.65 SD) to calculate an upper likely value for the error of a measuring system. This is like adding an apple and a pear. As we know from looking at QC results, individual results for the same material contain different magnitudes of error, which is why we see a dispersion of results. This is why the TE approach cannot be applied to individual results because it does not allow for the possibility that for any individual patient result the imprecision could differ from 1.65 SD.

In summary, although measurement error cannot be exactly known, the Total Error Concept is of theoretical value, and in practice is useful

in situations where setting an acceptable upper limit of total error for measurement results is required e.g. plasma cholesterol measurements, drugs in sport. The approach is not suitable when considering the uncertainty of individual results.

What is Measurement Uncertainty (MU)?

In contrast to TE, MU is not concerned with estimating measurement error. Routine laboratories generally measure a patient sample once rather than many times, and therefore the MU approach focuses on identifying the dispersion of results that might have been obtained for an analyte if a sample had been measured repeatedly instead of once. To do this, the MU approach uses available data about repeated measurements from a given measuring system to define an interval of values within which the true value of the measured analyte is believed to lie, with a stated level of confidence. For example, if a plasma glucose concentration result was 4.5 mmol/L, availability of appropriate MU information might give the laboratory ~95 % confidence that the true value of the glucose concentration in the sample lies in the range 4.4-4.6 mmol/L.

In summary, MU does not estimate error, but provides a quantitative estimate of where the true value of a measured analyte is believed by the laboratory to lie, with a stated confidence level. As such, the term measurement uncertainty tends to give the wrong impression, as it is actually a quantitative indication of the level of confidence, or belief, the laboratory has about the quality of a result. MU is therefore an essential parameter of the reliability of measurement results. The basic parameter of MU is 1 SD. Note: MU is a property of measurement results, not the measurement procedure producing them.

What does MU include?

The many potential sources of variability of patients measurement results are traditionally classified as pre-measurement (e.g. within person biological variation, stress, drugs, sample transport etc.) measurement and post-measurement (e.g. result rounding). For many analytes the magnitude of pre-measurement variabilities swamp those associated with the measurement process. However, MU is concerned only with the uncertainties sourced within the measuring process itself e.g. primary tube sampling or sample preparation to result output. For specific analytes or clinical purposes pre-analytical uncertainties, if expressible as an SD, can be combined with MU.

How does MU handle bias?

Bias is a predictable offset value relative to an appropriate

reference e.g. assigned value of a certified secondary or conventional reference material, or a peer group or all mean target value in an EQA or laboratory round-robin if improved inter-assay alignment is sought. Whatever approach is used to determine a bias value for a routine measurement procedure, the MU approach assumes that known bias is eliminated or minimised e.g. by re-calibration. As discussed above, a bias value cannot be exactly known, and therefore bias cannot be completely eliminated. The MU approach recognises that the value used for bias correction has an associated uncertainty, being the combination of the uncertainty of the reference value itself (if available), and the standard error of the mean value obtained from the replicate measurements of the reference produced by the measurement procedure. The uncertainty of the bias value is therefore expressed as an SD. If bias cannot be estimated, then its magnitude is unknown and it cannot be addressed.

In summary, if bias is known and considered significant, then it should be eliminated or minimised in the measuring system. As to whether a bias is considered significant might be determined statistically, or be a professional judgement eg. a bias of 0.1 mmol/L for plasma calcium measurements is obviously significant, whilst a bias of 0.001 mmol/L is unlikely to be considered so. If a bias value has been estimated, the uncertainty of the value used for bias correction has to be considered for inclusion in the calculation of the overall MU for results produced by a given measurement procedure.

How is MU estimated?

Since the MU approach requires that known bias is eliminated or minimised, or ignored if unknown, we are left with:

- imprecision of the measuring system
- imprecision of the bias value used if bias was eliminated or minimised.

For most measuring systems the magnitude of the intermediate reproducibility imprecision is large relative to that for bias correction, and therefore for most measurement procedures MU is simply the intermediate reproducibility assessed using QC, usually expressed as 1 SD or CV %.

Measurement procedures are often automated 'black box' systems that do not permit components (e.g. sampling/reagent probes, water-baths, spectrophotometers etc) to be individually studied to ascertain their uncertainty. Fortunately, we are interested in the combined effect of the individual sources of variability on measurement results, and this is adequately reflected in the dispersion of results obtained for QC samples. Since patients results are compared with each other and with

reference values over time, it is appropriate that the QC data used for MU calculations is obtained over a period of time sufficient to capture variability due to routinely occurring changes in the measuring system e.g. reagent and calibrator batch changes, different operators, routine maintenance etc. i.e. intermediate reproducibility. Most measurement procedures are sufficiently robust that imprecision generally changes little between reagent batches, so imprecision can be calculated by combining their SDs (see below on how to combine SDs). Adequate data (>100 results) takes longer to obtain for infrequently performed measurement procedures, in which case interim calculations are appropriate, but in any case, including new procedures, a minimum 30 QC results is required before an approximate Gaussian distribution of data points can be reasonably assumed. Thereafter, as QC results accumulate, the imprecision should be regularly re-calculated until the SD is stable at the same number of decimal places used for reported results.

At its simplest, the mean value and SD is calculated for each level of QC used for a given measurement procedure over a sufficient time to encompass as many routine procedure changes as possible; at least 30 values is be adequate for an initial MU estimate. The parameter of MU is 1 SD (*standard measurement uncertainty*, symbol μ). Because the SD of the QC reflects the combined effect of all the individual uncertainties arising within the measuring system, the SD can be considered as the combined standard uncertainty (μ_c) for patients results around the mean value of the particular QC.

Since ±1 SD covers only ~68 % of the dispersion of obtained QC values, the uncertainty is widened by applying a coverage factor (k) to provide an *expanded measurement uncertainty* (symbol *U*). Usually k = 2 is chosen, to provide a more useful 95.5 % coverage of the dispersion of results. Assuming such a dispersion also applies to patients results, then a result could be in the form x ± y (95 % confidence), where y = 2 SD (i.e. 2 x μ_c = *U*). If several levels of QC are used the MU should be calculated for each, and a judgement made as to whether they are sufficiently different to warrant their use with patient results that fall in the range considered to be covered by each QC level.

Consider the following QC data for the serum rhubarb measurement procedure. The calibrator values are assigned internally by the assay manufacturer, and SI-traceable or conventional reference materials are unavailable to enable bias assessment. Performance relative to the peer group in an external quality assessment program showed bias from the group mean was always <0.1 mmol/L. Given the clinical application of the results, bias was not considered significant, and therefore ignored for MU calculation.

QC	07/27/11 – 09/14/11	Mean (mmol/L)	SD (μ_c)	2 SD (U) ($U = 2 \times \mu_c$)
level 1	n=86	4.9	0.12	0.24 = 0.2
level 2	n=86	28.7	0.73	1.46= 1.5

The imprecision under intermediate reproducibility conditions is used for calculating MU. Since patient results are reported to one decimal place, the expanded MU (U) is similarly treated i.e. Patients' results in the range considered monitored by:

QC level 1: x1 ± 0.2 mmol/L (95.5 % confidence); QC Level 2: x2 ± 1.5 mmol/L (95.5 % confidence).

e.g. Patient 1 result: 6.3 ± 0.2 mmol/L (~95 % confidence). This means that the laboratory has ~95 % confidence that the true value lies in the range 6.1–6.5 mmol/L.

Patient 2 result: 34.6 ± 1.5 mmol/L (~95 % confidence). This means that the laboratory has ~95 % confidence that the true value lies in the range 33.1–36.1 mmol/L.

The best estimate of the true value is always the reported result, but this way of expressing MU indicates that other results could have been obtained. Note: error is not mentioned. MU is concerned with the probability of where a true value lies. Note: the term 'true value' is in relation to the reference used for calibration, and may be an arbitrarily set value e.g. WHO International Units.

What if Bias is considered significant?

Suppose rhubarb assay results are clinically interpreted relative to decision values defined by an international expert body, and incorrect interpretations may have deleterious medical implications. Measurement bias is therefore important. Fortunately the assay manufacturer assigns calibrator values using a certified secondary (matrix-matched) reference material (CRM) that is metrologically traceable to the SI unit (mole), and claims the calibrator is commutable with the CRM. The laboratory purchases a vial of CRM and measures it 10 times under repeatability conditions.

	Certified value	1 SD (μ_c)	Lab repeatability study (n=10)	1 SD (μ_c)	Bias
CRM	3.87 ± 0.028 mmol/L (95.5 % CI)	0.014 mmol/L	Mean = 3.97 ± 0.12 mmol/L (95.5 % CI)	0.06 mmol/L	0.10 mmol/L

The positive bias of 0.1 mmol/L is judged significant and applicable across the measuring range, so the assay is re-calibrated down by that value. We need to calculate the uncertainty of the value of 0.1 mmol/L. Such calculations use 1 SD (μ_c), not 2 SD (U) so the standard uncertainty of the CRM is 0.014 mmol/L. The uncertainty of the mean value obtained by the laboratory is the standard error of the mean of the ten measurements i.e. $0.06/\sqrt{10} = 0.019$ mmol/L. We now combine the uncertainties of the CRM and the laboratory mean values to give the combined standard uncertainty of the bias value of 0.1 mmol/L. Because SDs cannot be added together they need to be converted to variances (SD^2), which can be added. The combined variance is then converted back to a combined SD by taking the square root e.g. the combined standard uncertainty of the bias value ubias = $\sqrt{(0.014^2 + 0.019^2)} = 0.0236$ mmol/L.

The uncertainty of the bias value should then be compared with the long term QC imprecision e.g. 0.0236 is ~11 % of 0.22 mmol/L, and is considered borderline large enough to be included in the calculation of the total MU of the results produced by the rhubarb procedure. Bias uncertainty and QC imprecision is combined in the same way as above.

For values around QC Level 1: $\mu_c = \sqrt{(0.0236^2 + 0.12^2)} = 0.1223$ mmol/L

Level 2: $\mu_c = \sqrt{(0.0236^2 + 0.73^2)} = 0.730$ mmol/L

The expanded standard uncertainty (U) for QC Level 1 is 0.122 x 2 = 0.244 mmol/L, (rounded to 0.2 mmol/L); and for QC 2: 0.73 x 2 = 1.46 mmol/L (rounded to 1.5 mmol/L).

Patients' results in the range considered monitored by QC level 1: x ± 0.2 mmol/L (95.5 % confidence); QC Level 2: x ± 1.5 mmol/L (95.5 % confidence). Note: MU values are rounded to the same number of decimal places as used for reporting results.

In this example inclusion of bias uncertainty made no meaningful change to the expanded uncertainty as determined using just the long term QC imprecision data. For this reason laboratories often ignore the uncertainty of bias values if they are less than an arbitrary cut-off of 20-30 % of the intermediate imprecision.

How is MU calculated for a measurement calculated from several other results?

e.g. Anion Gap = (Na+ + K+) – (Cl- + HCO3-)

Same as above. Suppose the uc (1 SD) for Na+ = 1.1 mmol/L, K+ = 0.1 mmol/L, Cl- = 1.2 mmol/L and HCO3- = 0.8 mmol/L

$\mu AG = \sqrt{(1.1^2 + 0.1^2 + 1.2^2 + 0.8^2)} = 1.82$ mmol/L; $U = 3.64$ mmol/L. Appropriate rounding gives an expanded uncertainty of ± 4 mmol/L.

Note that although the AG calculation includes addition and subtraction, the standard uncertainties are combined in the same way. If a calculated parameter includes divisions and/or multiplications (e.g. creatinine clearance), then the SDs must first be converted to CV before calculation i.e. $\sqrt{CV1^2 + CV2^2 + CV3^2 + \ldots}$ etc).

How should MU estimates be assessed?

Before embarking on calculating MU, it is essential for a laboratory to set clinically acceptable MU targets for each analyte e.g. serum sodium, urine sodium etc. There is little point in estimating MU if there are no targets stating what is required for clinically acceptable performance. This important aspect will not be discussed here as approaches to target setting are well described elsewhere. e.g. use of biological variation data, international expert group recommendations, professional opinion.

Summary

The approach described above is referred to as 'top down'. Known significant bias should be eliminated or minimised, and residual bias assessed in terms of the uncertainty of the bias value used for re-calibration or result correction. Bias uncertainty is often trivial relative to imprecision and is ignored, so that intermediate QC imprecision data captures the overall uncertainty of measurement results. MU data should be periodically updated.

The 'bottom up' approach estimates the uncertainties associated with individual components of a measuring system, and combines them in a model to reflect their effect in the complete measuring system. This approach is best suited to the needs of IVD medical device manufacturers validating new measurement procedures or seeking technical steps where MU might be reduced, and also for labs developing in-house measurement procedures. For the bottom-up approach readers are referred to the recent CLSI C-51 guideline [3].

Should MU be routinely reported to clinicians?

No, but should be available if requested e.g. clinical trials, clinical research.

MU useful to the laboratory because it:

- provides quantitative evidence that measurement results meet

clinical requirements for reliability

- is essential for meaningful comparison of results with reference values, with previous results using the same measurement procedure*
- can provide insights as to which technical steps might be open to improvement, thereby reducing overall MU
- is an essential component for achieving standardised and harmonized measurement results through metrological traceability.

*MU can be used to assess whether a patient's result is measurably different with ~95 % confidence from a reference value, or from a previous result, or combined with intra-individual biological variation, in exactly the same way as described by Fraser (4).

Summary of MU and TE

- TE provides an approximate worst case value for the error of a measuring system.
- TE does not recognise that each individual patient result could have other possible outcomes with less error than Bias + 1.65 SD.
- TE is useful for setting upper limits of allowable error.
- MU is not concerned with estimating the total error of a measuring system
- MU is concerned with estimating an interval of values within which the 'true' value of a measured analyte is believed to lie, with a stated level of confidence.
- Known bias is eliminated or minimised
- MU considers a single measurement result to be the best estimate of a true value, and centres on it the dispersion of other values that could have been obtained if the measurement had been repeated (usually with ~95 % confidence).
- MU is the appropriate approach for meaningfully comparing measurement results with reference values and previous results of the same kind.

References

1. White GH. Basics of estimating measurement uncertainty. Clin Biochem Rev 2008;29:S53-S60.
2. Requirements for the estimation of measurement uncertainty. National Pathology Accreditation Advisory Council, Australian Government Department of Health and Ageing, 2007. www.health.gov.au Search for NPAAC publications (accessed 2/04/2012).
3. C51-A. Expression of Measurement Uncertainty in Laboratory Medicine. CLSI, Wayne, PA.
4. Fraser CG. Biological Variation: From principles to practice.

   2001; AACC Press, Washington DC.