Van Eenoo P, Spaerkeer A, Delbeke FT

Measurement Uncertainty: a practical top-down approach

Doping Control Laboratory, Ghent University (UGent), Technologiepark 30, B-9052 Zwijnaarde, Belgium.

1. Introduction

Measurement uncertainty (MU) is defined by Eurachem [1] and the international laboratory accreditation cooperation document on MU; ILAC-G17 [2], according to which measurement uncertainty is a parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand. Hence, MU is a term that is of utmost importance for the interpretation of analytical results. However, although the principles of MU are well established in the community of analytical chemists, the methods to derive this parameter are still a topic of debate [3, 4]. Indeed, it is important to note that the documents of both organisations state that this parameter is an estimate that characterises the range within which a measured value is asserted to lie with a certain probability. Therefore, since it only concerns an estimate, it should be noted that for MU there is - in contrast to other parameters analytical chemists deal with- NOT a true value.
And hence, several calculation methods and approaches to derive this estimate can be used. One approach is often called the bottom-up approach and uses statistical and metrological methods to incorporate all conceivable sources of uncertainty of a method [1]. After evaluating which steps might lead to significant contributions, this data is than processed mathematically to obtain measurement uncertainty.
The second approach is generally called top-down and starts from real data, which means results obtained from multiple measurements [1].
Since doping control most often deals with non-threshold substances, the statement in ILAC-G17 that “for now only MU in quantitative testing is considered” is extremely important [2]. Although MU is a well established concept in doping control laboratories for several years now, only limited data is available on the calculation of MU that is applicable to this field

[3, 5, 6]. In this paper the top-down approach at DoCoLab to estimate the uncertainty and evaluate this estimation is described.

2. Initial considerations and specificity of the doping control setting

MU statements are as fragile as the weakest link in their determination and therefore a laboratory needs to use instruments, standards, reagents, consumables and – of course last but not least- personnel that is fit for purpose.
Some practical examples are the preferred use of positive displacement rather than air displacements pipettes for stock solutions prepared in organic solvents, the use of instruments that are under continuous monitoring of the quality system and the use of adequate reference standards and solutions. Indeed, keep in mind that stock solutions degrade and preferably different solutions for qc's and calibrators should be used.

It is also generally accepted that uncertainty is concentration dependent. Therefore, if one wants to identify and state the uncertainty for every single analytical result there will be a need to derive MU for several points starting at the lowest and continuing up to the highest point of the calibration curve.
However, in doping control this is only of minor concern. Indeed, for doping control the situation can be reduced to an evaluation of compliancy/non-compliancy with the rules, meaning does the concentration exceed the threshold or not? [7, 8]
Indeed, it is not required to report the uncertainty for each result, but to allow the responsible authorities to take a correct decision within a 95% probability setting [9].
If measurement uncertainty is reduced to this type of compliancy evaluation there are several consequences. Indeed, in such cases there is a need to use one-sided statistics and importantly MU can not be attributed to the individual result which means in practice no plus/minus statement, but only a statement of compliance or non-compliance with the rules.

3. Example

An example is our method for the quantitative detection of the main metabolite of cannabis with a 5 point calibration curve to determine the concentration of this substance [10]. Each sample is always analyzed in triplicate and accompanied by a blank urine sample, a water sample and a quality control sample spiked at the threshold. This positive control sample is

spiked and analyzed independently from the calibration samples and contains the major metabolite of cannabis at the threshold concentration.
The method itself consists out of carefully pipetting 2.0 ml urine, an alkaline hydrolysis at 60 degrees for 10 minutes, followed by a double liquid-liquid extraction step with a mixture of n-hexane and ethyl acetate after acidification of the urine with glacial acetic acid.
Finally the combined extracts are evaporated, TMS derivatised and analyzed by GC-MS in the SIM-mode.
Therefore there is a clear need to identify the relationship between the measurand and the input parameters [1]. This is often done using a cause – effect diagram which looks like a fish bone skeleton (Figure 1).

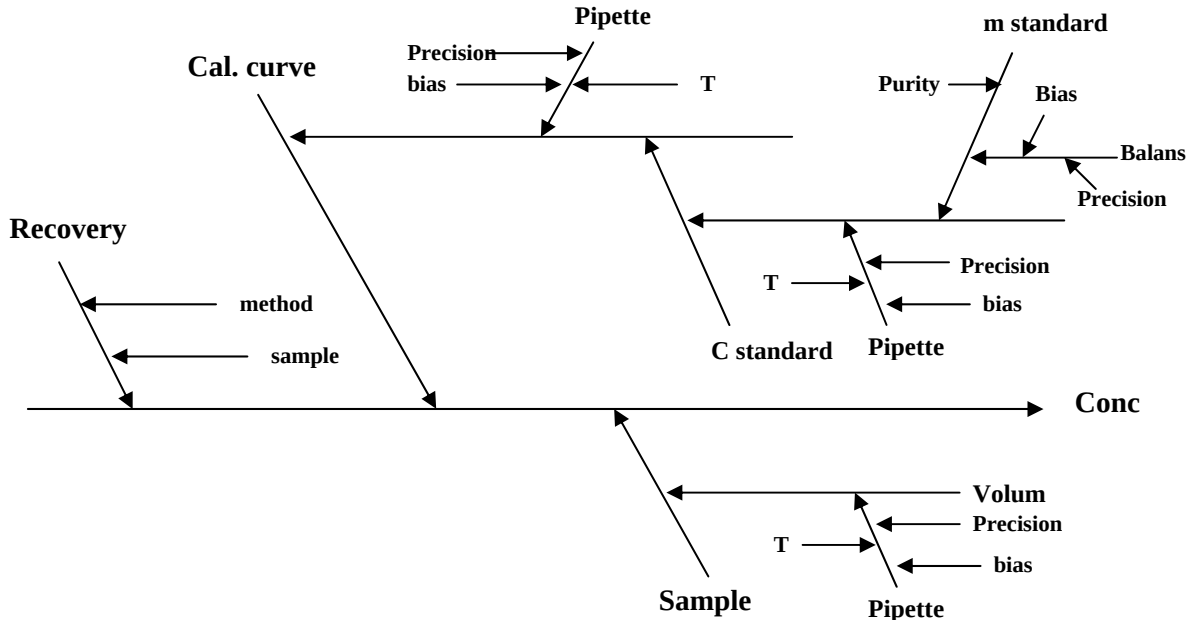

Figure 1. Cause-effect diagram of a quantitative method using liquid-liquid extraction at DoCoLab.

The results of all a positive control samples is processed in a quality control chart. Over a period of several months or even years a huge amount of information is gathered in this way. And because of the extended time period, this data is obtained under different conditions. Indeed, over such a long period the data was surely obtained using different calibration curves, technicians as well as different lots of reagents, consumables, instruments and even standards or stock solutions of standards. Of course this means that the cause effect diagram can be simplified by incorporating all individual sources that can affect precision into one contributor, precision of the measured QC-values.
Similarly also a part of the influence of what was identified as the recovery contribution is covered, since over a longer period of time also different blank urine samples, representing

realistic changes in matrix that will lead to small differences in recovery will be used as well as different other factors influencing the recovery including reagents and instrumental settings (for example the real temperature of the ovens for hydrolysis and derivatisation) will be covered, as long as these parameters are under stringent control of a quality system like ISO17025.

In this way, it is quite easy to get the most realistic estimate of long term precision of the method incorporating all possible contributing sources for this factor.

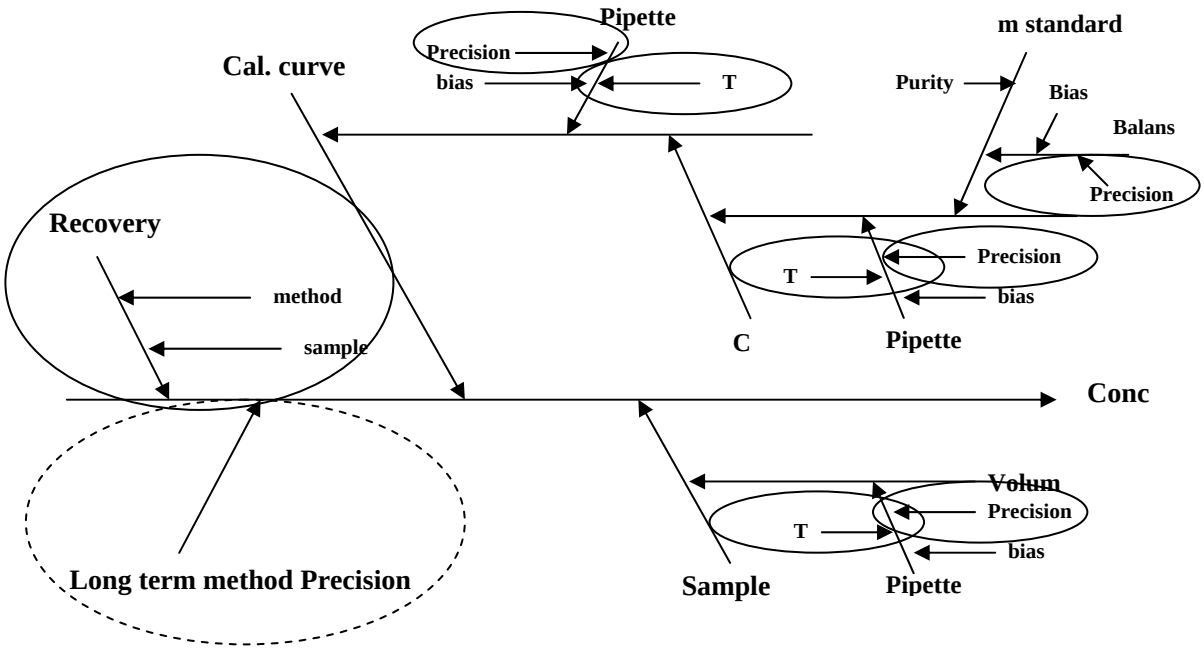

Figure 2. Effect on the cause-effect diagram through incorporation of the long-term precision data obtained from a QC-chart.

Using the same quality control charts obtained for precision data, the long term bias of the method can be estimated**.** This long term bias is the difference between the theoretical concentration at which the QC-samples were spiked and the mean of the measured values. Preferably such data is based upon QC-samples at the threshold from a certified reference urine. If this is not available, a QC-sample spiked with a different source of reference standard can act as a surrogate, at least the stock solution for the QC and calibration samples should be from a different preparation..

Again, when this is done over a long period of time, a realistic idea of the contributing factors is obtained and the cause-effect diagram can be simplified, incorporating all bias factors, including those in the recovery, into one log-term bias parameter, as shown in Figure 3.

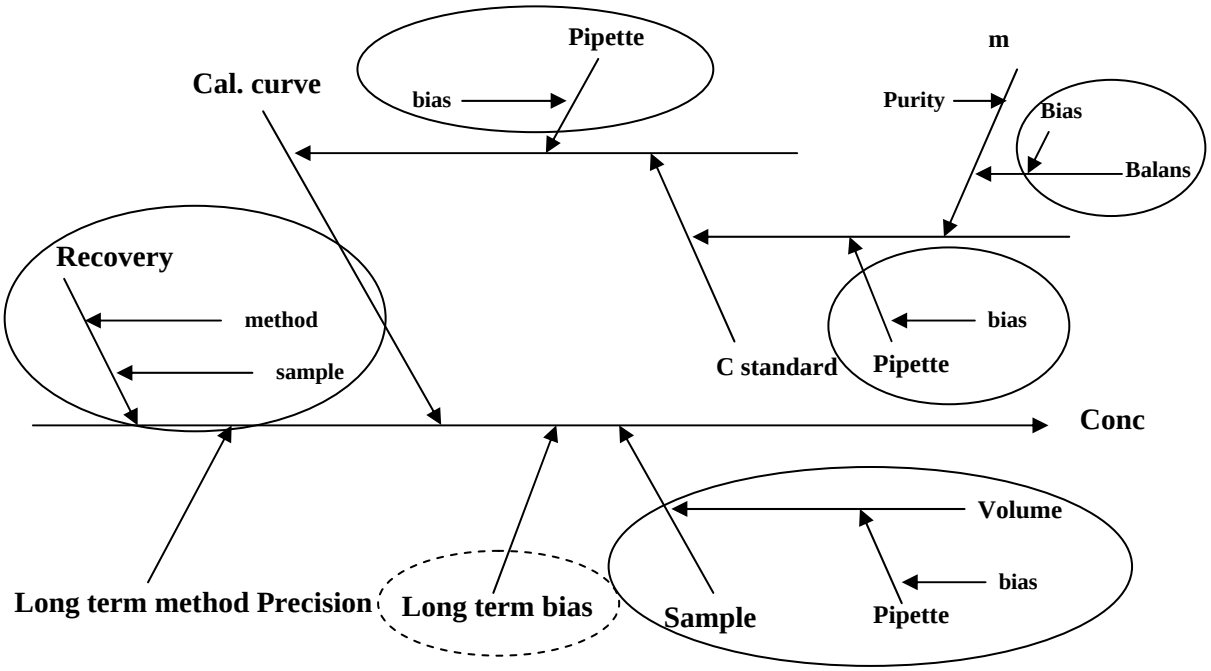

Figure 3. Effect on the cause-effect diagram through incorporation of the long-term bias data obtained from a QC-chart.

Hence, using data readily at hand for existing methods, the cause-effect diagram can be drastically reduced to three major contributors: long-term precision, long-term bias and the purity of the reference standard (Figure 4).

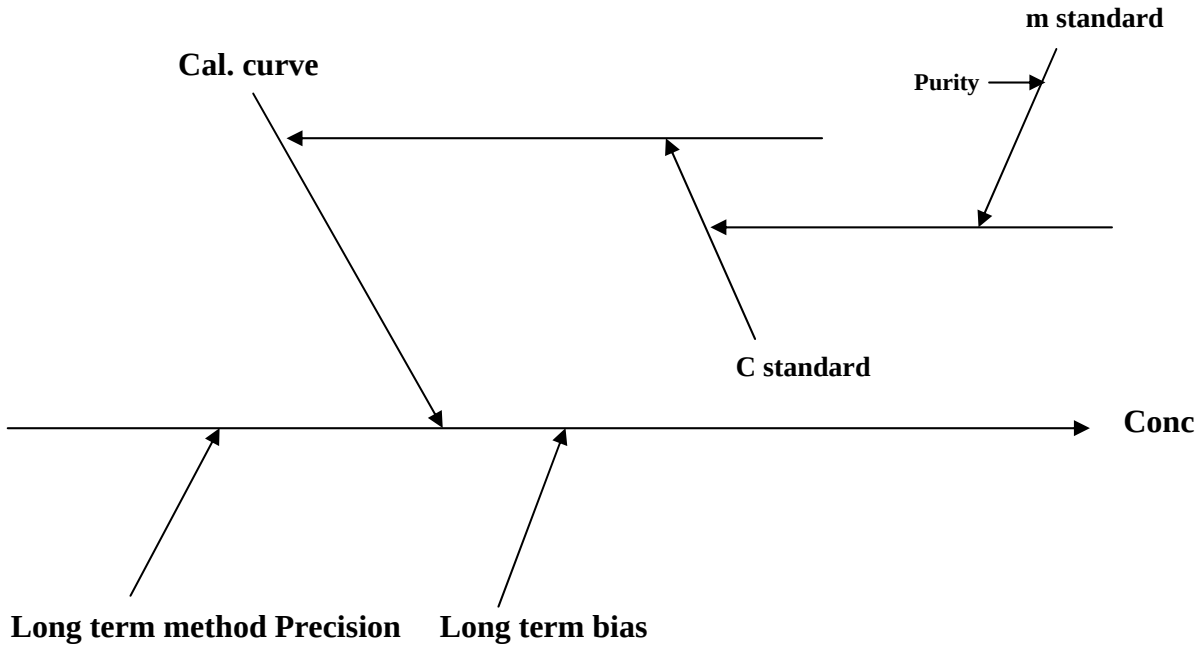

Figure 4. Final cause-effect diagram after incorporation of the long-term bias and precision data obtained from a QC-chart

Since the data on the quality of reference standards is mentioned on the certificate of analysis, all contributors to uncertainty are known and the standard uncertainty of the method can be calculated via:

$$u = \sqrt{(U_{precision})^2 + (U_{bias})^2 + (U_{purity})^2}$$

The expanded uncertainty U is calculated by multiplying the standard uncertainty u by a factor k.

Taking into account that in compliancy testing one sided statistics is needed and that WADA's ISL requests a 95%-confidence level, k=1.64 assuming a Gaussian distribution [8]. Based upon these results, decision limits can be calculated above which a sample is identified as non-compliant, i.e. the concentration exceeds the threshold.

This needs to be done by adding the expanded uncertainty to the threshold concentration.

4. Overview of MU at DoCoLab and the influence of the contributors

It has been suggested that the top-down approach often only includes precision data and that this leads to an underestimate of the measurement uncertainty. In this method bias is incorporated as well and the methodology offers a good opportunity to evaluate the significance of all incorporated parameters (precision, bias and reference standard purity). Table 1 shows each of the contributors for the quantitative methods used at DoCoLab.

Table 1. Threshold substances, threshold and uncertainty contributors

substance	threshold	u precision	u bias	u purity	u
epitestosterone	200 ng/ml	12,14	15,8	0,008	19,9
cafeine	12 µg/ml	0,664	0,11	0,00003	0,673
T/E	4	0,23	0,5		0,55
cathine	5 µg/ml	0,465	0,289	0,00003	0,547
phenyl propanolamine	25 µg/ml	2	0,933	0,00004	2,497
ephedrine	10 µg/ml	0,983	0,55	0,0005	1,126
pseudoephedrine	25 µg/ml	2,281	1,095	0,0007	2,53
methylephedrine	10 µg/ml	0,568	0,164	0,00011	0,591
THC-COOH	15 ng/ml	1,298	1,078	0,0004	1,687
morphine	1 µg/ml	0,113	0,022	0,00001	0,115
salbutamol	500 ng/ml	36,5	33,8	0,007	49,7
norandrosterone	2 ng/ml	0,195	0,022	0,00006	0,196

As shown in Table 1, the purity of the reference standard is never a significant contributor. However, the uncertainty component of the bias . Sometimes it is even the major contributor, e.g. the testosterone to epitestosterone ratio.
In case of THC-COOH, the u is 1.687 ng/ml or U = k x u= 2.77 ng/ml (for k=1.64) [8].

5. Evaluation of MU estimations

Because MU-determinations are estimations, there is a clear need to evaluate the obtained estimations for correctness. Results of proficiency tests can offer such possibility.
It can be assumed that the relative uncertainty in a small concentration range does not change significantly. Hence, if the concentrations of threshold substances in proficiency testing samples are similar to the threshold concentrations, this data can be used to evaluate the MU estimation.
Therefore the dispersion of the results of an individual laboratory should be in agreement with the true/consensus value of a PT test, taking into account MU. Hence, the % deviation of the individual result should be smaller than the expanded uncertainty expressed as a percentage.
In case of THC-COOH, the calculated expanded uncertainty of 2.77 ng/ml represents 18.47%.

true value	lab	deviation as percentage
12,2	13,1	7,38
19,9	19,5	2,01
34,1	32,7	4,11
30,4	28,5	6,25
27,4	25,6	6,57
24,8	23,7	4,44
23,1	20,7	10,39
19	18,9	0,53
29,9	30,6	2,34
Th	U @ 95%	
15	2,77	18,47

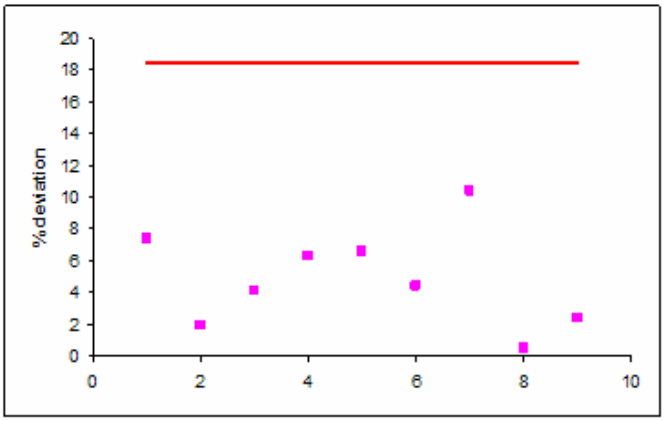

Figure 4. PT-results and deviation between the reported value by DoCoLab and the true value of the PT-sample, expressed as a % and represented in a table as well as graphically.

The evaluation of the MU estimation can then be achieved easily through inspection of the graphic representation of the results. Indeed, from Figure 4 it is clear that 95 percent of the obtained results lie beneath the boundary set by MU. Within this area there also needs to be a

realistic spread that does not suggest that the uncertainty is either overestimated (all deviations are far from maximum allowed deviation) or underestimated (all deviations are close to the maximum). In the example shown (Figure 4), such a spread is present.

Finally, it is clear that MU is not a static concept and evaluations and calculations of MU needs to be performed periodically and updated regularly.

6. Acknowledgements

The postdoctoral grant by the Flemish Ministry of Culture, Youth, Sports and Brussels (PVE) is gratefully acknowledged.

7. References

1. Eurachem/CITAC Guide (2000) Quantifying Uncertainty in Analytical Measurement. Second Edition.
2. International Laboratory Accreditation Cooperation (ILAC) (2002). ILAC G17:2002 Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025.
3. Van Eenoo, P., Delbeke F.T. (2003). Reply to “Measurement uncertainty and doping control in sport” by A. van Der Veen, Accred Qual Assur (2003) 8:334-339. *Accred Qual Assur* **8**, 334-339.
4. van der Veen, A.H. (2003). Measurement uncertainty and doping control in sport *Accred Qual Assur* **8**, 477-479.
5. Ventura, R., Ventura Garcia, M., , Bocqué, R., Maroto, A. (2003). Estimating uncertainties of analytical results using information of internal quality control, intercomparison exercises and reference materials. In: Schänzer, W., Geyer, H., Gotzmann, A., Mareck, U. (eds.) *Recent Advances in Doping Analysis (11)*, Köln, pp. 131-140.
6. Leung, G.N.W., Ho, E.N.M., Kwok, W.H., Leung, D.K.K., Tang, F.P.W., Wan, T.S.M., Wong, A.S.Y., Wong, C.H.F., Wong J.K.Y., Yu, N.H. (2007). A bottom-up approach in estimating the measurement uncertainty and other important considerations for quantitative analyses in drug testing for horses. *J. Chromatograph A* **1163,** 237-246
7. Faber, N.M., Boqué, R. (2006). On the calculation of decision limits in doping control. sport *Accred Qual Assur* **11**, 536-538.
8. Eurachem/CITAC Guide (2007) Use of Uncertainty in formation in compliance assessment. First edition.
9. World Anti-Doping Agency (2004). The World Anti-Doping Code. International Standard for laboratories 4.0.
10. De Cock, K.J.S., Delbeke, F.T., De Boer, D., Van Eenoo, P., Roels, K. (2003). Quantitation of 11-nor-tetrahydrocannabinol-9-carboxylic acid with GC-MS in urine collected for doping analysis. *J. Anal. Toxicol.* **27**, 106-109.