

# Determination of the standard deviation for proficiency assessment from past participant's performances

Isabelle Côté · Piotr Robouch · Benjamin Robouch ·  
David Bisson · Philippe Gamache · Alain LeBlanc ·  
Pierre Dumas · Mikaël Pedneault

Received: 30 January 2012 / Accepted: 30 May 2012 / Published online: 19 June 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** The “uncertainty function” introduced by Thompson et al. estimates the reproducibility standard deviation as a function of concentration or mass fraction. This model was successfully applied to data derived from three proficiency testing schemes aiming at the quantification of cadmium, lead and mercury in blood and urine. This model allows the estimation of standard deviation for the performance assessment for proficiency testing rounds.

**Keywords** Horwitz equation · Thompson-modified equation · Uncertainty function · Limit of detection · Reproducibility · Proficiency testing

## Introduction

The “Centre de toxicologie du Québec” (CTQ) [1] belonging to the “Institut national de santé publique du Québec” (INSPQ) is a public organization that has been offering human toxicology expertise (environmental, clinical and

occupational) to the provincial health network of Quebec (Canada) as well as to external clients from around the world. Since 1979, the CTQ operates several permanent external quality assessment schemes that enable participating laboratories to evaluate the accuracy and precision of their analytical methods on a continuous basis. Approximately 250 laboratories from over 30 countries participate in these proficiency testing (PT) schemes to analyse a wide variety of elements in biological PT materials of human origin, such as blood, serum, urine or hair.

In order to verify and further confirm the applicability of the “uncertainty function” described and discussed in several publications [2–8], we compiled all the reference values ( $X_{\text{Ref}}$ ) and the corresponding reproducibility standard deviations ( $s_{\text{R}}$ ) determined in the frame of three PT programs designed for the determination among others of three toxic trace elements (cadmium, mercury and lead) in blood and urine matrices. A total of 861 data pairs ( $X_{\text{Ref}}$ ,  $s_{\text{R}}$ )—later denoted as the CTQ data—were analysed to identify whether similar trends are observed in the seventeen cases under investigation (3 PT schemes; 3 elements; 2 matrices; one PT scheme does not monitor mercury in urine).

Presented at the Eurachem PT Workshop, October 2011, Istanbul, Turkey.

I. Côté (✉) · D. Bisson · P. Gamache · A. LeBlanc ·  
P. Dumas · M. Pedneault  
Centre de toxicologie du Québec (CTQ)/INSPQ,  
945, ave Wolfe, Québec, QC G1V 5B3, Canada  
e-mail: isabelle.cote@inspq.qc.ca

P. Robouch  
European Commission, Joint Research Centre,  
Institute for Reference Materials and Measurements (IRMM),  
Retieseweg, 111 Geel, Belgium

B. Robouch  
Istituto Nazionale di Fisica Nucleare (INFN),  
Laboratori Nazionali di Frascati, Via E. Fermi 40, Frascati, Italy

## Methodology

The experimental data evaluated in this work were reported in the frame of the three PT programs described hereafter:

- The “Interlaboratory Comparison Program for metals in biological matrices” (PCI) is a bimonthly scheme attended by over 130 laboratories applying their routine analytical techniques;
- The “Priority Metals Quality Assessment Scheme” (PMQAS) is a scheme designed for the US State

**Table 1** Summary of the CTQ data, including the number of data points ( $N$ ), the mass fraction ranges investigated ( $C$ ), the fitted parameters  $\alpha$ ,  $\beta$  and calculated ratios  $\alpha/\beta$ , for cadmium, mercury and

lead in urine and blood, assigned in the frame of the PCI, QMEQAS and PMQAS proficiency testing schemes. The relative standard errors of the parameters and their ratio are indicated between parentheses

Element	Matrix	PT	$N$	$C$ (g g <sup>-1</sup> )	$\alpha$ (g g <sup>-1</sup> )	$\beta$	$\alpha/\beta$ (g g <sup>-1</sup> )
Cd	Blood	PCI	63	(0.11–1.5)×10 <sup>-8</sup>	2.0×10 <sup>-10</sup> (30 %)	0.084 (17 %)	2.4×10 <sup>-9</sup> (34 %)
		PMQAS	75	(0.06–7.9)×10 <sup>-8</sup>	1.5×10 <sup>-10</sup> (27 %)	0.054 (15 %)	2.8×10 <sup>-9</sup> (31 %)
		QMEQAS	22	(0.05–1.3)×10 <sup>-8</sup>	1.5×10 <sup>-10</sup> (36 %)	0.088 (22 %)	1.7×10 <sup>-9</sup> (42 %)
	Urine	PCI	63	(0.06–1.6)×10 <sup>-8</sup>	1.5×10 <sup>-10</sup> (45 %)	0.075 (22 %)	2.0×10 <sup>-9</sup> (50 %)
		PMQAS	60	(0.02–1.8)×10 <sup>-8</sup>	2.8×10 <sup>-11</sup> (41 %)	0.047 (19 %)	6.0×10 <sup>-10</sup> (45 %)
		QMEQAS	24	(0.14–1.3)×10 <sup>-8</sup>	1.9×10 <sup>-10</sup> (45 %)	0.075 (18 %)	2.5×10 <sup>-9</sup> (48 %)
Hg	Blood	PCI	63	(0.18–8.3)×10 <sup>-8</sup>	6.4×10 <sup>-10</sup> (27 %)	0.104 (20 %)	
		PMQAS	75	(0.01–1.2)×10 <sup>-7</sup>	3.6×10 <sup>-10</sup> (30 %)	0.052 (18 %)	6.9×10 <sup>-9</sup> (35 %)
		QMEQAS	22	(0.15–6.1)×10 <sup>-8</sup>	6.7×10 <sup>-10</sup> (37 %)	0.080 (17 %)	8.4×10 <sup>-9</sup> (41 %)
	Urine	PCI	63	(0.02–2.6)×10 <sup>-7</sup>	8.5×10 <sup>-10</sup> (30 %)	0.110 (19 %)	7.7×10 <sup>-9</sup> (36 %)
		PMQAS		No data			
		QMEQAS	24	(0.40–9.2)×10 <sup>-8</sup>	1.4×10 <sup>-9</sup> (33 %)	0.150 (17 %)	9.3×10 <sup>-9</sup> (37 %)
Pb	Blood	PCI	63	(0.17–8.4)×10 <sup>-7</sup>	3.4×10 <sup>-9</sup> (33 %)	0.067 (17 %)	5.1×10 <sup>-8</sup> (37 %)
		PMQAS	75	(0.02–1.4)×10 <sup>-6</sup>	3.0×10 <sup>-9</sup> (23 %)	0.049 (21 %)	6.1×10 <sup>-8</sup> (31 %)
		QMEQAS	22	(0.24–5.6)×10 <sup>-7</sup>	2.4×10 <sup>-9</sup> (31 %)	0.058 (23 %)	4.1×10 <sup>-8</sup> (39 %)
	Urine	PCI	63	(0.12–7.7)×10 <sup>-7</sup>	2.8×10 <sup>-9</sup> (28 %)	0.076 (16 %)	3.7×10 <sup>-8</sup> (32 %)
		PMQAS	60	(0.004–4)×10 <sup>-7</sup>	–	0.038 (22 %)	
		QMEQAS	24	(0.02–1.0)×10 <sup>-6</sup>	2.7×10 <sup>-9</sup> (23 %)	0.063 (14 %)	4.3×10 <sup>-8</sup> (27 %)

Laboratories, all equipped with the same experimental instrumentation (inductive coupled plasma mass spectrometry, ICP-MS) and applying the same experimental protocols for the analysis of trace elements in blood and urine; and

- The “Quebec Multi-element External Quality Assessment Scheme” (QMEQAS) attended by 60 laboratories using ICP-MS.

At the end of the each PT round, a classical statistical treatment was applied to the results reported by the participants to calculate—after outlier rejection—the median value and the reproducibility standard deviation ( $s_R$ ). The median value was set as the assigned reference value ( $X_{Ref}$ ), while  $s_R$  was used to derive the standard deviation for performance assessment ( $\sigma_{PT}$ ). The CTQ data were compiled from the previous PT rounds organized by the CTQ for cadmium, mercury and lead in blood and urine matrices, as indicated in Table 1. All values were systematically converted to mass fraction (g g<sup>-1</sup>).

In the early 1980s, Horwitz et al. [9] reviewed the reported results in the frame of the Association of Official Analytical Chemists (AOAC) PT rounds and derived an empirical relation estimating the coefficient of variation for the reproducibility ( $CV_R$ ) as a function of the mass fraction ( $C$ ) expressed in g g<sup>-1</sup>:

$$CV_R = 2^{(1-0.5 \lg C)} / 100 \quad (1)$$

Twenty years later, Thompson re-evaluated [10] the results reported in several PT schemes and confirmed the validity of the Horwitz equation at mass fractions ranging from 1.2×10<sup>-7</sup> to 0.138 g g<sup>-1</sup>, while suggesting a constant  $CV_R$  of 0.22 (or 22 %) for mass fractions below 1.2×10<sup>-7</sup> g g<sup>-1</sup>.

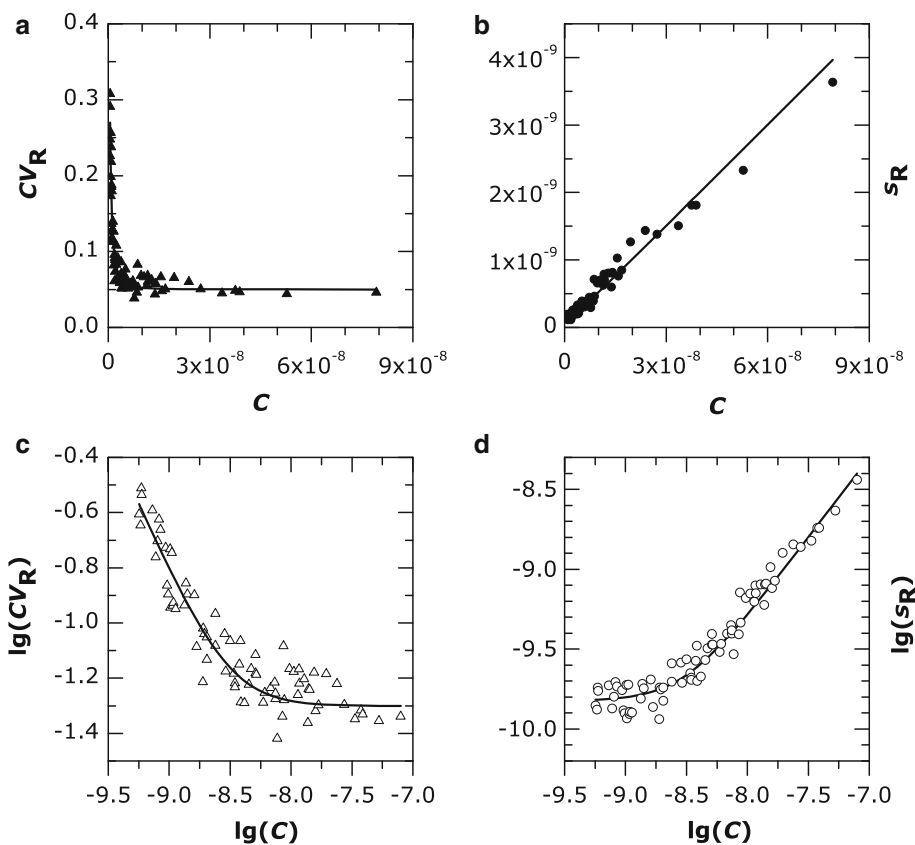
In order to have a clearer view on how to proceed, we plotted a set of four graphs for each combination of PT scheme, element and matrix, namely:

- $CV_R$  versus  $C$ , as suggested by Horwitz [9];
- $s_R$  versus  $C$ ;
- $\lg(CV_R)$  versus  $\lg(C)$ , where “lg” denotes the logarithm to base 10; and
- $\lg(s_R)$  versus  $\lg(C)$ , as suggested by Thompson [7, 10].

An example of such a set of graphs is shown in Fig. 1a–d presenting all the data collected in the frame of the PMQAS round for the determination of cadmium in blood. A constant  $CV_R$  of approximately 0.05 is observed at higher concentration (Fig. 1a)—equivalent to a linear increase in  $s_R$  with increasing  $C$  (Fig. 1b)—while a constant  $s_R$  is observed at the lowest mass fraction range (Fig. 1d). These observations are consistent with the “uncertainty function” introduced by Thompson in 1988 [2] and further discussed since [3–7]:

$$s_R = \sqrt{\alpha^2 + \beta^2 C^2} \quad (2)$$

**Fig. 1** The four graphical representations of the CTQ data for Cd in blood obtained in the frame of the PMQAS proficiency testing scheme: **a**  $CV_R$  versus  $C$ ; **b**  $s_R$  versus  $C$ ; **c**  $\lg(CV_R)$  versus  $\lg(C)$ ; and **d**  $\lg(s_R)$  versus  $\lg(C)$ . The *solid line* represents the “uncertainty function” fitting the experimental data points.  $C$  and  $s_R$  are expressed in  $\text{g g}^{-1}$

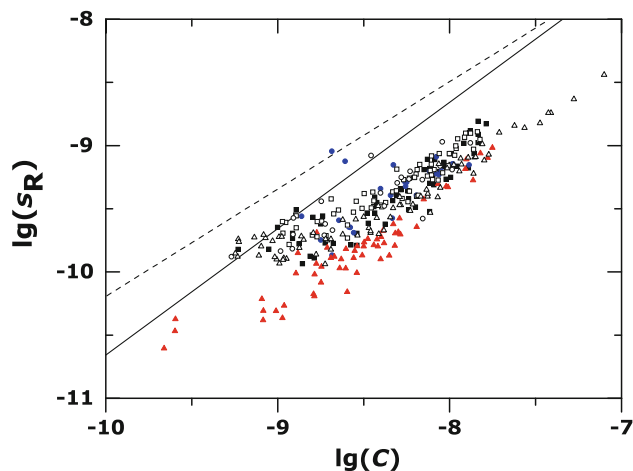


The function described in Eq. 2 was systematically fitted to the CTQ data. The Newton-Raphson algorithm implemented in the Microsoft Excel 2010 Solver was used—without any further data weighting—to minimize the sum of squares of residuals and to derive the two parameters  $\alpha$  and  $\beta$ . The initial value of parameter  $\beta$  was set equal to the  $CV_R$  of the highest mass fraction investigated, while the initial value of parameter  $\alpha$  was set to  $10^{-9}$  by default.

## Results and discussion

Neither the Horwitz model nor the Thompson-modified one fits the CTQ experimental data obtained for cadmium, mercury or lead. The example for cadmium illustrated in Fig. 2 clearly shows that most of the data lie below the two model curves.

On the other hand, the “uncertainty function” (Eq. 2) fits well the CTQ data. An example is shown in Fig. 1. As stated by Thompson [8], the “uncertainty function” is function of parameter  $\alpha$  “[...] describing the constant variation at concentrations close to the detection limit [...]” and of parameter  $\beta$  representing “[...] the constant relative standard deviation at high concentration [...]”. On the basis of this assumption, an alternative mathematical



**Fig. 2** The CTQ data for cadmium in urine (*filled symbols*) and blood (*empty symbols*) assigned in the frame of the PCI (*squares*), QMEQAS (*circles*) and PMQAS (*triangles*) proficiency testing schemes. Most of the data lie below the Horwitz (*dashed line*) and the Thompson-modified (*solid line*) model curves.  $C$  and  $s_R$  are expressed in  $\text{g g}^{-1}$

approach was derived from Eq. 2, confirmed the values obtained for  $\alpha$  and  $\beta$ , and allowed the estimation of the respective relative standard errors from the corresponding variations, which could not be obtained using the MS Excel 2010 Solver. For each PT-matrix-element combination, an estimate of  $\beta$  was calculated as the average of  $CV_R$  at the

high concentration range. The mean value of  $\alpha$  was then derived as:

$$\alpha = \sqrt{s_R^2 - \beta^2 C^2} \quad (3)$$

When plotting the “uncertainty function” versus mass fraction, one gets the characteristic shape predicted by Horwitz [9]—sometimes referred as the “Horwitz trumpet” [11]—and described by Thompson [2]. Figure 1c and d shows that the “uncertainty function” has two asymptotes—represented by a constant reproducibility standard deviation, below a certain mass fraction; while above it, represented by a constant coefficient of variation of the reproducibility. The two asymptotes intercept at a mass fraction equal to the ratio  $\alpha/\beta$ . Table 1 presents the mass fraction ranges, the values for  $\alpha$ ,  $\beta$  and the ratio  $\alpha/\beta$ , together with the respective relative standard errors—provided between parentheses—for the seventeen PT-matrix-element combinations investigated.

Reliable  $\beta$  values are determined with a relative standard error ranging from 14 to 23 %. The  $\beta$  values for the PMQAS program are systematically the smallest of the order of 0.05, as expected from a PT scheme having participants using the same experimental protocol and the same instrumentation. The other PT schemes display  $\beta$  values of 0.07, 0.08 and 0.11 for Pb, Cd and Hg, respectively (Table 1). Koch and Magnusson reported similar results [12].

Assuming that the “uncertainty function” remains applicable down to mass fraction close to the limit of quantification ( $C_{LOQ}$ ) and to the limit of detection ( $C_{LOD}$ ) one could estimate following indicative upper limits:

$$C_{LOD} = 3\alpha \text{ or } C_{LOQ} = 10\alpha \quad (4)$$

Fewer and more scattered data were available for the determination of  $\alpha$ , for which the relative standard errors ranged from 23 to 45 %.  $\alpha$  values of 0.2, 1 and 3  $\mu\text{g kg}^{-1}$  were obtained for Cd, Hg and Pb, respectively (Table 1). This would correspond to estimated limits of detection of 0.6, 3 and 9  $\mu\text{g kg}^{-1}$  in blood and urine matrices. These limits are well above—up to 20 times—those determined experimentally for a specific sample treatment and a dedicated instrumental technique. Such over-estimated values may be due to the fact that the presented  $\alpha$  values derive from reproducibility standard deviations (computed from results reported in the frame of several PT schemes, and obtained using various analytical methods), whereas  $C_{LOD}$  are usually determined under repeatability conditions.

The ratio  $\alpha/\beta$  for each element from the different matrices and PT schemes are in agreement within 20 %, when excluding the value for the PMQAS Cd in urine. Ratios of the order of 2, 8 and 48  $\mu\text{g kg}^{-1}$  were obtained

for Cd, Hg and Pb, respectively (Table 1). When combining with Eq. 4, the following approximations are derived:  $\alpha/\beta \approx 2C_{LOQ}$  for  $\beta = 0.05$  (i.e. PMQAS) or  $\alpha/\beta \approx C_{LOQ}$  for  $\beta = 0.10$  (i.e. PCI or QMEQAS). This indicates that CTQ might have organized some PT rounds close to the limit of quantification, below which measurement relative uncertainties higher than 22 % are to be expected. This could explain the high scatter of data points at the low concentration range.

## Conclusion

The “uncertainty function” introduced by Thompson et al. [2] describes well the trend of reproducibility standard deviation versus mass fraction for cadmium, mercury and lead in blood and urine samples. The compilation of  $\alpha$  and  $\beta$  calculated using a simple mathematical approach (without any data weighting) will allow CTQ to estimate reproducibility standards deviations and to derive the standard deviation of performance assessment ( $\sigma_{PT}$ ) for various PT-element-matrix combination. The robust statistical treatment prescribed by the ISO 13528 guide [13] would be performed for confirmation. On the other hand, participants could use the same function to calculate the reproducibility standard deviation to derive a reasonable estimate of their measurement uncertainty, as prescribed by the Eurolab [14]. The CTQ intends to evaluate the “uncertainty function” for the remaining elements and matrices available.

Furthermore, the CTQ will evaluate the dispatch of identical PT samples in several PT schemes, using, for example, the assigned values ( $X_{Ref}$ ,  $\sigma_{PT}$ ) of one PT round to the other PTs, similar to what is implemented by the International Measurement Evaluation Program (IMEP) [15]. This could significantly reduce the costs for homogeneity and stability investigation, ensuring the propagation of sound metrological principle to various groups of participants.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. <http://www.inspq.qc.ca/ctq/paqe> (Accessed April 11th, 2012)
2. Thompson M (1988) Variation of precision with concentration in an analytical system. *Analyst* 113:1579–1587
3. Thompson M (1998) Perspective: do we really need detection limits? *Analyst* 123:405–407
4. Thompson M, Wood R (2006) Using uncertainty functions to predict and specify the performance of analytical methods. *Accred Qual Assur* 10:471–478

5. Thompson M, Mathieson K, Damant AP, Wood R (2008) A general model for interlaboratory precision accounts for statistics from proficiency testing in food analysis. *Accred Qual Assur* 13:223–230
6. Jiménez-Chacón J, Alvarez-Prieto M (2009) Modeling uncertainty in a concentration range. *Accred Qual Assur* 14:15–27
7. Thompson M (2011) Uncertainty functions, a compact way of summarising or specifying the behaviour of analytical systems. *Trend Anal Chem* 30:1168–1175
8. Thompson M, Coles BJ (2011) Use of the ‘characteristic function’ for modelling repeatability precision. *Accred Qual Assur* 16:13–19
9. Horwitz W, Kamps LR, Boyer KW (1980) Quality assurance in the analysis of foods for trace constituents. *J Assoc Off Anal Chem* 63:1344–1354
10. Thompson M (2000) Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. *Analyst* 125:385–386
11. Eppe G, Van Cleuvenbergen R, Samstuen Haug L, Boulanger B, Becher G, De Pauw E (2008) Empirical relationship between precision and ultra-trace concentrations of PCDD/Fs and dioxin-like PCBs in biological matrices. *Chemosphere* 71:379–387
12. Koch M, Magnusson B (2012) Use of characteristic functions derived from proficiency testing data to evaluate measurement uncertainties. *Accred Qual Assur*. doi:10.1007/s00769-012-0880-8
13. ISO 13528 (2005) Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO, Geneva
14. Eurolab Technical Report No.1/2007 ([www.eurolab.org](http://www.eurolab.org))
15. Duta S, Taylor P, Boiciuc D, Iacobescu F (2008) Improving metrology in chemistry infrastructure in Romania via IRMM-JRC-EC collaboration. *Metrologie* 2–3:13–20