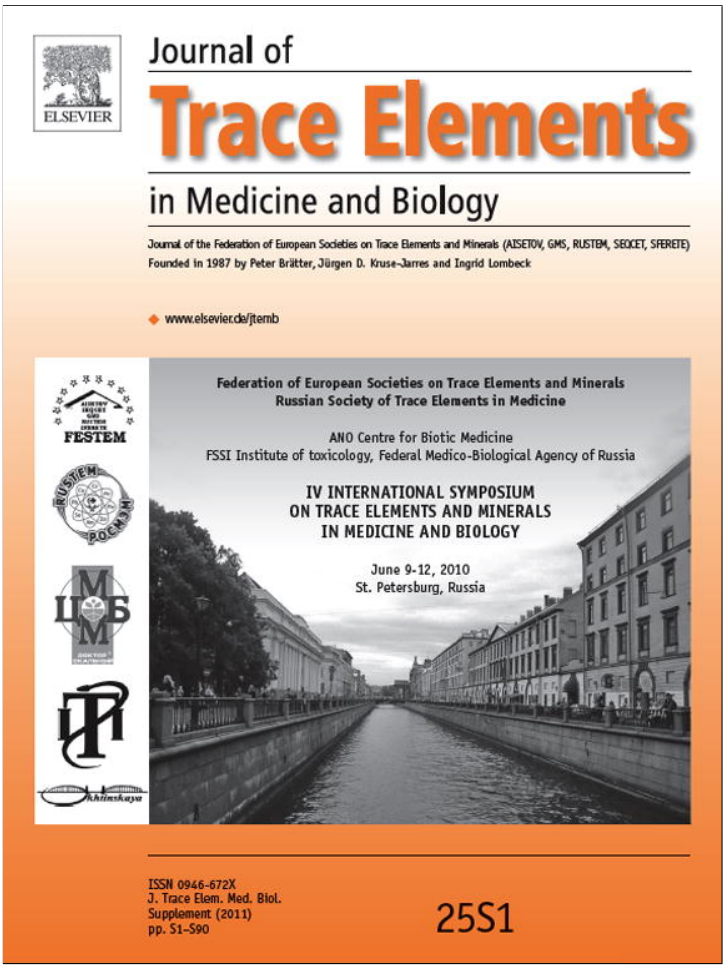

Journal of Trace Elements in Medicine and Biology 25S (2011) S17–S21

Contents lists available at ScienceDirect

Journal of Trace Elements in Medicine and Biology

journal homepage: www.elsevier.de/jtemb

FOURTH INTERNATIONAL FESTEM SYMPOSIUM

Quality assessment of measurement☆

Andrew Taylor[a,b,*]

[a] Trace Element Centre, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK
[b] Department of Clinical Biochemistry, Royal Surrey County Hospital, Guildford, UK

ARTICLE INFO

Article history:
Received 2 July 2010
Accepted 1 November 2010

Keywords:
Trace element
Quality assurance
Quality specifications
External quality assessment

ABSTRACT

Assessing the quality of measurements is of interest to organizers of external quality assessment schemes (EQAS, or proficiency testing schemes), laboratory analysts and managers, users of laboratory results and other agencies. Scheme organizers run test programmes, define standards of acceptable and non-acceptable performance, and interact with participants and oversight authorities. Laboratory personnel are responsible for the quality management system and to choose whether to accept the standards set by scheme organizers or to adopt their own. Users receive and act upon the laboratory results.

Schemes within the same analytical sector are often organized very differently causing contradictory assessment of performance. The Network of EQAS in occupational and environmental laboratory medicine established collaborative projects designed to enhance assessment of measurement quality and to improve the reliability of laboratory results.

To address the issue of variations in assessing the quality of measurements, and in response to comments from some participants, standards derived from biological variation, rather than on the analytical performance of participants have been developed. Evaluation of test materials with respect to homogeneity and stability, and work on methods to give the assigned value to test materials, has also been undertaken. Following from these projects, further collaboration is planned which will provide better quality assessment of measurement to scheme participants and their users.

Introduction

Quality assessment is a necessary feature in any analytical laboratory. But, it is important to define what is meant by *quality assessment of measurement* and who has responsibility for this activity.

The *measurement* is a procedure carried out in the laboratory to answer a particular question. Usually this is on behalf of a 'client', such as a paediatrician who needs to know, for example, "does this child with abdominal pain have lead poisoning?" or "what are the blood lead concentrations in this group of children with epilepsy?"

Reference to *quality* indicates the need for a measurement to be fit for purpose. Will the question be answered satisfactorily? If the limit of detection of the assay used to measure the concentration of lead in blood is 10 μg/dL, will toxicity be detected, will the blood lead concentrations in the epileptic children be defined accurately? The answer would be 'yes' to the first of these challenges, as concentrations of lead in symptomatic poisoning will be much higher than 10 μg/dL, but 'no' to the second where concentrations will be lower than this value. Quality here implies the existence of quality standards or quality specifications.

Procedures to determine whether a result meets the quality specification, and the method is fit for the purpose for which it is being used, are part of the *assessment*.

Responsibility for quality assessment of measurement is shared by a number of interested parties; laboratory personnel, legislative agencies, accreditation bodies, metrologists, and organizers of external quality assessment (EQA) schemes (or proficiency testing schemes).

Laboratory personnel: Managers and analytical staff will define the quality specifications appropriate for the work they undertake. This includes determination of the required accuracy, precision and detection limits for an assay. They will also develop and monitor internal quality control procedures and react to any issues identified. Responding to external prompts is a further way in which personnel exercise responsibility for quality assessment of their measurements. This could be the actions taken when a report is received from an EQA scheme, or when comments or complaints are received from users of the results and who have reason to suspect that an analysis is failing to meet the quality specification.

In certain sectors, quality standards are recommended or even imposed on the laboratory by regulatory bodies or interested professional associations [1,2]. For example, national authorities

☆ Presented at the 4th International Symposium on Trace Elements and Minerals in Medicine and Biology, 9–12 June 2010, Saint-Petersburg, Russia.

* Correspondence address: Trace Element Centre, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK.
E-mail addresses: a.taylor@surrey.ac.uk, andrewtaylor4@nhs.net

0946-672X/$ – see front matter © 2010 Elsevier GmbH. All rights reserved.
doi:10.1016/j.jtemb.2010.11.001

Table 1
Differences in statistical results produced by different EQA schemes when examining the same set of data. Results are in μmol/L [7].

Scheme number	Consensus mean	CV (%)	Number of outliers
1	0.74	17	2
2	0.71	8	9
3	0.73	16	3
4	0.72	13	6
5	0.74	16	4
6	0.73	27	2

with oversight responsibility for occupational monitoring demand minimum performance standards from laboratories carrying out measurements, and in some countries the standards are a legal requirement. Measurement of aluminium in serum to protect patients receiving haemodialysis against the risk of aluminium toxicity is facilitated by standards proposed by renal physicians. Statistical expertise as applied to data handling for EQA schemes is provided by metrologists and those with related interests [1,3–6].

The most important and objective assessment of the quality of measurements is afforded by participation in EQA schemes [5]. Organizers design and run assessment programmes, and this will include developing their own quality specifications against which participants are rated as satisfactory or unsatisfactory. Parameters that are used when determining scheme specifications include the analytical state-of-the-art, the views of participants, the views of users of laboratory results and of legislators. Scheme organizers take one or more of these parameters into account when forming their quality specifications.

Thus, quality assessment of measurement involves a wide range of interested groups and individuals.

External quality assessment schemes

While EQA schemes set out to assess the quality of measurements, there are huge differences in the way in which schemes are organized even within the same analytical sector. The number and frequency of challenges vary, the test items may be formulated from human, animal or even artificial sources, there is little consistency among the statistical routines used to examine results and the criteria to define satisfactory and unsatisfactory performance have little in common. Consequently, when different scheme organizers process the same set of results different, and in some cases, conflicting outcomes are produced.

This was first demonstrated by Christensen and Olsen [7] when the same set of blood lead results were examined and interpreted by six EQA schemes. As is shown in Table 1, the tests used to identify and eliminate outlier values produce different numbers and, therefore, the consensus data are also different. Organizers were asked to indicate which laboratories gave satisfactory or unsatisfactory results according to the criteria used for their scheme, and to provide a ranking order. One scheme had no criteria and was unable to comply; the responses of the other five are shown in Table 2. While most schemes were in reasonable agreement about the best and worst participants there was considerable variation in ranking the mainstream laboratories. Of even more concern was that a laboratory could be deemed to be satisfactory in some schemes and unsatisfactory in others – on the basis of the same set of results.

Scoring the quality of measurement performance

Schemes almost always score measurement performance following calculation of the proximity of a result to the assigned value;

$$x - X$$

Table 2
Ranking and performance rating by 5 EQA schemes working with the same data sets. Scheme 2 rates performance as either satisfactory (Y) or not satisfactory (N). Scheme 4 includes a group with performance that is neither satisfactory nor unsatisfactory. The scoring system used by Scheme 5 does not provide for subtle discrimination. Scores of 8–10 are regarded as satisfactory. Satisfactory and unsatisfactory participants are indicated by (*) or (**), respectively [7].

Lab no.	Scheme number				
	1	2	3	4	5
57	1	Y*	1	2*	10*
96	2	Y*	5	1*	10*
09	3	Y*	2	7*	10*
36	4	Y*	3	6*	10*
32	5	Y*	9	5*	8*
94	6	Y*	6	2*	8*
25	7	Y*	4	8*	10*
83	8	N**	20	13	7**
28	9	Y*	12	12	7**
41	10	N**	10	11	9*
03	11	N**	8	17	8*
06	12	N**	23	22**	5**
67	13	N**	22	21**	5**
90	14	N**	7	4*	8*
71	15	N**	11	20**	9*
40	18	N**	13	15	7**
59	20	N**	17	9	8*
87	23	N**	16	10	8*
39	28	N**	29	29**	1**

where x = laboratory result and X = the assigned value (often the consensus mean).

The internationally recommended assessment involves calculation of a z-score [5] where the proximity to the assigned value is compared against a measure of the variance at the concentration of the assigned value:

$$z\text{-score} = \frac{x - X}{s}$$

where 's' represents the variance at the assigned value, sometimes described as the "standard deviation for proficiency testing".

The conventions employed state that satisfactory and unsatisfactory performance are indicated by a z-score of 2 or less, or greater than 3, respectively. However, if schemes use different values for 's', the ambiguities demonstrated in Table 2 will continue to emerge. Therefore, together with colleagues in the Network of Organizers of EQA Schemes for Occupational and Environmental Laboratory Medicine, we have attempted to introduce consistency by developing analytical quality specifications using objective data, and applying these to the z-score calculation.

Analytical quality specifications

The intention which was set out when developing these quality specifications was to demonstrate that the measurement allowed the detection of a significant change in the analyte concentration within a subject, i.e. that the measurement is fit for purpose. This objective was approached by determining the total allowable error (*TEa%*) as described by Fraser [1].

$$TEa\% < Bias\% + z \times \mathrm{CV}_a\%$$

where

$$Bias\% = 0.25 \times \sqrt{\mathrm{CV}^2_{\mathrm{intra}} + \mathrm{CV}^2_{\mathrm{inter}}}$$

and

$$\mathrm{CV}_a\% = 0.5 \times \mathrm{CV}_{\mathrm{intra}}$$

Table 3
Quality specifications formulated from the total allowable error formula of Fraser [1].

Assay	Quality specification
Lead in blood	±3 μg/dL or ±10%, whichever is the greater
Aluminium in serum	±5 μg/L or ±20%, whichever is the greater
Copper in serum	±0.84 μmol/L or 12%, whichever is the greater
Selenium in serum	±0.072 μmol/L or 12%, whichever is the greater
Zinc in serum	±1.20 μmol/L or 15%, whichever is the greater

and CV_{intra} and CV_{inter} refer to intra- and inter-individual variability for a given parameter and $z = 1.65$ for a 95% probability level.

Using recently published data, together with some of our own investigations we have prepared quality specifications for 5 elements, as given in Table 3 [8,9]. These provide for up to a fixed percentage deviation from the assigned value over most of the concentration range of interest, but, to allow for the increase in imprecision at low analyte concentrations there is then a numerical component. As shown in Fig. 1, the specifications then conform to a 'funnel' shape.

By applying these quality specifications to the *z*-score formula, assessment of performance will be consistent among schemes and will also indicate whether or not a measurement procedure is giving results that are fit for the purpose for which the method is being used;

$$z\text{-score} = \frac{x - X}{s/2}$$

the quality specification for the *TEa* is divided by two to conform with the ISO requirement that a *z*-score of <2 shall indicate satisfactory performance.

For example, if a blood sample has a concentration of 40 μg/dL, any result within ±10% ($s = 4$ μg/dL) would be regarded as satisfactory and laboratory reporting a result of 44 μg/dL would achieve a *z*-score of +2

The assigned value

As can be seen from Table 1, a further source of variation among schemes can be given by selection of the assigned value. ISO Standard 17043 [5], and related documents, describe five ways to set assigned values, i.e.

1. By formulation (i.e. by spiking known amounts)
2. By certification (analysis using methods of high metrological order)
3. By prior analysis together with certified reference materials (CRMs), where the CRM is used as the calibrant
4. By consensus of results from expert laboratories (using robust statistics)
5. By consensus of results from all participants (using robust statistics)

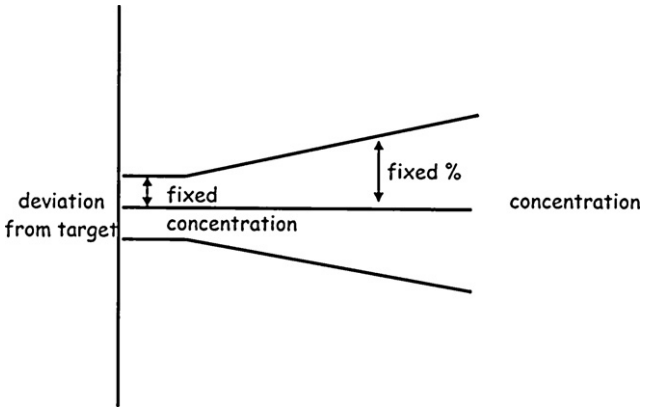

Fig. 1. Graphical representation of the relationship between analyte concentration and quality specification.

To evaluate the reliability of these procedures in the context of EQA schemes for trace elements in clinical samples a comparison experiment was carried out by the Network.

Three samples were prepared from the same pool of human serum, one was unspiked and the other two were made with accurately known additions of copper, selenium and zinc. The pools were dispensed into aliquots for analysis as follows:

- Measurement by isotope-dilution inductively coupled plasma (ID-ICP-MS) at a national metrological laboratory
- Measurement by a small group of reference laboratories, together with a serum CRM
- Measurement by a small group of reference laboratories, to give their consensus data
- Measurement by approximately 300 laboratories who were participants in EQA schemes organized in Belgium, Canada, France, Italy, Netherlands, Spain, UK and USA, to give their consensus data

The results are shown in Table 4. The consensus data were calculated using Algorithm A from ISO 13528 [6]. While the variancies, as indicated by the expanded uncertainty and the standard deviations, increased as the procedures include more variables, there was very little difference between the actual assigned values. Furthermore, there was no consistency to suggest that one procedure is always closer to the results given by ID-ICP-MS than others. The conclusion from this investigation is that the mean of the results from all participants, calculated using robust statistics, may be taken as the assigned value.

However, when the participant's means are determined for the individual schemes, Table 5, it is evident that this procedure can give misleading information. For reliable estimate of the assigned value using the consensus mean a minimum number of results must be reported. This number will vary depending on the analyte and sample type and scheme organizers should carry out investigations to ensure that, if they adopt this approach, participants can have confidence in the assigned value.

Other contributions to quality assessment of measurement

While the primary aim of EQA schemes is to provide for quality assessment of the measurements made by participants, schemes contribute to assessments in other ways. We have shown that for measurements that are difficult to perform and present a substantial challenge to the analyst, such as the determination of aluminium, laboratories that have a special interest in trace elements perform better that those where the assays form a minor component of the test repertoire and workload [10].

External quality assessment schemes may also be designed to demonstrate whether particular analytical methods are suitable for use. Colorimetric methods to measure copper and zinc are being used increasingly, particularly where they may be adapted for the multi-channel clinical analysers in large hospital laboratories. However, results given by these methods are far from reliable. Table 6 shows that more than 20% of participants in the UK EQA scheme employ colorimetric methods to measure the concentration of zinc in serum with a smaller number also using kits to determine copper. All of the participants deemed to have

Table 4
Assigned values for copper, selenium and zinc in three samples of serum. Results are given as μmol/L. The values reported for isotope dilution inductively coupled plasma-mass spectrometry (ID-ICP-MS) are given with the expanded uncertainty (*U*, *k* = 2). Other values are the mean and standard deviation (SD). CRM = certified reference material.

	Copper	Low	Medium	High
ID-ICP-MS	Result	13.62	21.50	29.37
	U	0.17	0.22	0.35
Assay with CRM	Result	14.26	20.77	28.14
	SD	0.24	0.24	0.50
Consensus data from reference laboratories	Result	13.61	20.89	27.83
	SD	0.85	1.33	0.71
Consensus data from scheme participants	Result	13.81	21.29	29.21
	SD	1.89	1.64	2.15
	Selenium	Low	Medium	High
ID-ICP-MS	Result	0.75	1.54	3.17
	U	0.028	0.056	0.120
Assay with CRM	Result	0.72	1.46	2.99
	SD	0.04	0.06	0.18
Consensus data from reference laboratories	Result	0.69	1.43	2.96
	SD	0.062	0.206	0.583
Consensus data from scheme participants	Result	0.74	1.52	3.10
	SD	0.10	0.15	0.36
	Zinc	Low	Medium	High
ID-ICP-MS	Result	9.89	23.00	31.50
	U	0.49	0.339	1.21
Assay with CRM	Result	10.99	24.14	31.67
	SD	0.40	0.40	0.46
Consensus data from reference laboratories	Result	9.46	22.61	29.94
	SD	0.45	1.15	1.38
Consensus data from scheme participants	Result	9.95	23.26	31.06
	SD	0.84	1.66	2.07

Table 5
Consensus mean values for results reported to the individual EQA schemes. The results reported by each participant for the three samples analysed were averaged and these values used to calculate the mean values shown. Results for selenium are not given as there were too few from some schemes.

	Copper				Zinc			
	n	Mean (μmol/L)	CV (%)	Outliers (%)	*n*	Mean (μmol/L)	CV (%)	Outliers (%)
Belgium	20	26.14	7.5	0	20	21.87	5.1	5.0
Canada	48	21.57	4.4	4.2	47	21.87	4.5	2.1
France	32	21.73	7.3	3.1	31	20.95	8.3	6.5
Italy	19	19.69	10.2	5.3	22	20.64	9.1	9.1
Netherlands	23	22.83	9.1	8.7	25	22.02	7.2	4.0
Spain	19	20.16	11.2	15.8	19	21.56	10.0	15.8
UK	79	22.20	6.9	7.6	95	21.25	10.1	3.2
USA	24	21.26	7.0	4.2	32	21.10	10.1	6.3
All labs	266	21.57	7.1	6.0	293	21.41	7.6	4.1

Table 6
Numbers of participants and poor performers in the UK NEQAS for trace elements, shown for all participants and for those using colorimetric methods.

	Number of participants		Poor Performers	
	All	Colorimetric	All	Colorimetric
Copper	83	6	0	0
Zinc	90	21	5	5

unsatisfactory performance for their zinc assays use a colorimetric method whereas results obtained by atomic spectrometric techniques are seen to be fit for purpose. Colorimetric assays for copper in serum pose fewer problems.

Summary

Quality assessment of measurement involves a range of individuals and groups, all of whom wish to be assured that results reported by a laboratory are fit for the purpose for which the measurement was undertaken. External quality assessment schemes have a major role to play in assessing the quality of measurement but the differences in the way in which schemes are organized, and handle data, can be so varied that they lead to contradictions in performance assessment.

The key to achieving equivalence in performance assessment among schemes lies in the standard deviation for proficiency testing, '*s*' as used to calculate a *z*-score. We have shown that an objective approach to deriving suitable '*s*' values is possible using the formula for total allowable error (*TEa*) to calculate quality specifications. If these quality specifications are used, equivalence in assessment among different schemes is given and assessment can also be linked to fitness for purpose. Fitness for purpose may be applied to analytical performance of individual participants and also to methodologies.

The assigned value for EQA scheme test samples may be derived from participant results, using robust statistics, but it is necessary to ensure sufficient number of results to avoid the introduction of a bias.

Acknowledgements

Sarah-Jane Bainbridge and Chris Harrington who prepare and distribute thousands of test samples each year, and generate reports for participants, at the UK NEQAS for Trace Elements, Guildford. Josiane Arnaud, Robert Jones, Mi-Young Lee, Olav Mazarrasa, Patrick Parsons, Marina Patriarca, Sinikka Valkonen, Jean-Philippe Weber, Cas Weykamp who have been responsible for the success of the Network of EQAS Organizers for Occupational and Environmental Laboratory Medicine.

References

[1] Fraser CG. General strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999;59:487–90.
[2] Kenny D, Fraser CG, Hyltoft P, Kallner A. Consensus agreement. Scand J Clin Lab Invest 1999;59:585.
[3] ISO 17025. General requirements for the competence of testing and calibration laboratories; 1999.
[4] ISO 15189. Medical laboratories – particular requirements for quality and competence; 2003.
[5] ISO 17043. Conformity assessment – general requirements for proficiency testing; 2010.
[6] ISO 13528. Statistical methods for use in proficiency testing by interlaboratory comparisons; 2005.
[7] Christensen JM, Olsen E. Comparison of some European external quality assessment schemes in the field of occupational and environmental medicine. Ann Ist Super Sanita 1996;32:285–94.
[8] Taylor A, Angerer J, Claeys F, Kristiansen J, Mazarrasa O, Menditto A, et al. Comparison of procedures used to evaluate laboratory performance in external quality assessment schemes for lead in blood and aluminum in serum demonstrates the need for common quality specifications. Clin Chem 2002;48:2000–7.
[9] Arnaud J, Weber J-P, Weykamp C, Parsons PJ, Angerer J, Mairiaux E, et al. Quality specifications for the determination of copper, zinc, and selenium in human serum or plasma: evaluation of an approach based on biological and analytical variation. Clin Chem 2008;54:1892–9.
[10] Taylor A, Walker AW. Measurement of aluminium in clinical samples. Ann Clin Biochem 1992;29:377–89.