Accred Qual Assur (2006) 11: 440–445
DOI 10.1007/s00769-006-0118-8

GENERAL PAPER

Andrew Taylor
Jurgen Angerer
Josiane Arnaud
Françoise Claeys
Robert L. Jones
Olav Mazarrasa
Eric Mairiaux
Antonio Menditto
Patrick J. Parsons
Marina Patriarca
Alain Pineau
Sinikka Valkonen
Jean-Philippe Weber
Cas Weykamp

Quality specifications for evaluation and comparison of performance among external quality assessment schemes in occupational and environmental laboratory medicine

Received: 31 October 2005
Accepted: 21 February 2006
Published online: 31 March 2006

Presented at the Eurachem PT Workshop September 2005, Portorož, Slovenia

A. Taylor (✉)
Centre for Clinical Science and Measurement, School of Biomedical and Molecular Sciences, University of Surrey,
Guildford, GU2 7XH, UK
e-mail: a.taylor@surrey.ac.uk
Tel.: +44-1483-689978
Fax: +44-1483-689979

J. Angerer
Institute of Occupational, Social and Environmental Medicine, University of Erlangen-Nuernberg,
91054 Erlangen, Germany

J. Arnaud
Département de Biologie Intégrée (DBI)-CHU de Grenoble,
BP 217 38043, Grenoble Cedex 9,
France

F. Claeys · E. Mairiaux
Unit of Epidemiology, Scientific Institute of Public Health,
1050 Brussels, Belgium

R. L. Jones
Nutritional Biochemistry, Division of Laboratory Sciences, National Center for Environmental Health, CDC,
Atlanta, GA 30341-3724, USA

O. Mazarrasa
Laboratorio de Higiene Industrial, Centro de Seguridad y Salud en el Trabajo,
Gobierno de Cantabria,
39012 Santander, Spain

A. Menditto · M. Patriarca
Department of Food Safety and Veterinary Public Health, Istituto Superiore di Sanità,
00161 Rome, Italy

P. J. Parsons
New York State Department of Health, Wadsworth Center Laboratories,
PO Box 509, Albany, NY 12201-0509,
USA

A. Pineau
Laboratoire de Toxicologie, UFR de Pharmacie, Université de Nantes,
44035 Nantes, France

S. Valkonen
Biomonitoring Laboratory, Department of Toxicology and Industrial Hygiene, Finnish Institute of Occupational Health,
00250 Helsinki, Finland

J.-P. Weber
Centre de Toxicologie, Institut National de Santé Publique du Québec,
945 Wolfe Avenue, Québec, Canada,
G1V 5B3

C. Weykamp
MCA Laboratory, Queen Beatrix Hospital,
7101 BN Winterswijk, The Netherlands

Abstract Quality specifications (QS) are proposed for lead in blood and for aluminium, copper, selenium and zinc in serum as part of the aim to set standards of performance for laboratories so that results can be demonstrated to be fit for the purpose to which they are applied. The QS were established taking account of the analytical state-of-the-art, physiological variations in the concentrations of the analyte and the clinical purpose for which the assay is to be used. A procedure was devised that uses these QS to give equivalence of assessment among external quality assessment schemes (EQAS), thus avoiding conflicting information which has been demonstrated in the past. Advantages of this procedure are: to provide direct comparison of performance of laboratories taking part in different schemes, to provide equivalence of assessment of laboratory performance necessary to establish mutual recognition agreements, and to demonstrate the fitness for purpose of results from participants.

Keywords Quality specifications · Occupational and environmental laboratory medicine · Z-score · Equivalence of performance assessment · Fitness-for-purpose

Introduction

As part of the aim to develop the scientific role of external quality assessment schemes (EQAS), one objective of the network of EQAS organisers in the field of occupational and environmental laboratory medicine (OELM) is to set standards of performance for laboratories, so that results can be demonstrated to be fit for the purpose

to which they are applied. A second objective is to develop procedures that permit equivalence of assessment among schemes, so that performance of laboratories taking part in different EQAS can be directly compared. To meet these aims and objectives, analytical quality specifications based on determinations of total allowable error (TEa) [1] have been developed and applied to performance monitoring.

The protocol of Kenny et al. [2] proposes that an analytical quality specification may be set after taking account of available information covering the analytical state-of-the-art, physiological variations in the concentrations of the analyte and the clinical purpose for which the assay is to be used. They recommend a hierarchical approach in which analytical data are at the lowest level and clinical information (opinions of clinicians, data relating to biological variability, and outcomes in specific clinical settings) is of greater importance. An approach has been developed to set clinically derived quality specifications, for which a general model described by Fraser [1], using data on biological variability to define total allowable error (TEa), can be employed. Such quality specifications may be used by analytical laboratories when evaluating their own methods. They may also be used by EQAS organisers when setting the denominator for the calculation of Z-scores, and thus provide for equivalence of performance assessment [3].

Organisers of EQAS use different statistical procedures to assess the performance of participants. This need not be a problem, provided that the outcomes are consistent among schemes. To test whether such consistency exists, in 1995 the organisers of five EQAS operating in the technical sector of OELM sought to compare how their schemes assess performance of participants. Thirty-two laboratories, selected because of their previous good performance record within the schemes, measured the concentration of lead in the same five samples of blood. The results obtained were then analysed by each of the EQAS organisers according to their usual procedures. This project showed that an individual laboratory's performance could be evaluated as unsatisfactory by one scheme but acceptable by another [4]. The same conclusion has since been found in similar projects that examined EQAS for haematology [5] and for analyses in water, food, soil and occupational hygiene [6]. It is apparent that inconsistency in performance assessment is a general problem.

According to international documents [7–9] for the organisation of EQAS (or proficiency testing, PT, schemes), Z-scores, or variants that involve the measurement uncertainty associated with the assigned value and/or the participant's result, can be used for performance scoring. Z-scores and its variants require the definition of a denominator to represent the allowed variability of results at the concentration of the assigned value of the test item. The importance of relating the denominator to fitness for purpose criteria was recently emphasised in the IUPAC revised document [9] which used the term 'σ_p' to express this concept. Even when schemes do use Z-scores, there are several ways as to how this denominator can be derived. Examples in use include the standard deviation (SD) of all results reported on a sample, the SD of results given by a group of expert laboratories, a percentage of the assigned value, a value that will include/exclude a given proportion of the participants. The IUPAC revised document now recommends using a value for σ_p that is deemed to produce a performance score to demonstrate whether a laboratory provides results that are fit for the intended use [9]. However, unless organisers of EQAS within the same technical sector collaborate and agree on the 'σ_p' value for application to their data, the performance scores will fail to provide comparable information and it will not be possible to show equivalence in the evaluation of laboratory performance in different schemes, nor for EQAS data to be taken into consideration when mutual recognition agreements are established among accreditation bodies and other organisations.

To introduce harmonisation in a way that will avoid imposition of a procedure from one scheme onto others and will also ensure that assessments demonstrate when analytical data from participants are fit for purpose, a procedure has been devised, which uses quality specifications, to give equivalence of assessment among EQAS in OELM.

Methods

Quality specifications

1. Data on current analytical performance were obtained from the results reported by participants of the EQAS of the network members. Table 1 briefly summarises the participant database of the schemes to give an indication of the numbers of results used in the following calculations and the analytical methodologies employed. These numbers are indicative as not every laboratory reported results on every occasion and not all schemes provided results for all elements.
2. Publications from organisations and associations with an interest in occupational and environmental health, toxicology, nutrition, dietetics and the clinical importance of essential trace elements were examined to determine whether any had produced recommendations on the analytical requirements to meet clinical decision making.
3. Following the work of Fraser [1], desirable targets for analytical imprecision (CV_a), bias and total allowable error (TEa) derived from biological variability may be expressed by the following formulae:

$$CV_a\% = 0.5 \times CV_{\text{intra}}$$

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

$$TEa\% < Bias\% + z \times CV_a\%$$

Table 1 Summary of participation numbers and analytical methodologies for Pb in blood and Al, Cu, Se, Zn in serum within the schemes organised by members of the network (2005)

	Blood Pb	Serum			
		Al	Cu	Se	Zn
Number of schemes	9	8	8	8	8
Total number of participants	420	250	350	180	400
Technique and percentage of results	%	%	%	%	%
FAAS	0	0	50	0	56
ETAAS	85	89	21	49	2
ICPMS	13	9	16	32	18
ICPAES	0	2	7	0	7
Colorimetry	0	0	6	0	17
Other	2	0	0	9	0

Abbreviations: *FAAS* flame atomic absorption spectrometry; *ETAAS* electrothermal atomic absorption spectrometry; *ICPMS* inductively coupled plasma mass spectrometry; *ICPAES* inductively coupled plasma atomic emission spectrometry

where 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. Similar formulae describe minimal and optimal targets. Publications reporting investigations of intra- and inter-individual variations for each of the elements included in this study were examined to provide data that could be used to calculate total allowable error.

Assessment of laboratory performance

In earlier work [3], six EQAS which included assessment of lead in blood within the menu of tests available to participants provided information from distributions in which the assigned values were approximately 100 μg/L (0.5 μmol/L). The data given by scheme organisers were:

- all reported results (x),
- the assigned value, following exclusion of outliers (X). (Each scheme has its own procedure to define outliers.)
- the differences between the reported results and the assigned value (x–X),
- the appropriate quality specification (QS), as given in Table 5, for the concentration of the assigned value.

The Z-score for each reported result was calculated as

$$z = \frac{x - X}{QS/2}$$

where the quality specification for TEa was divided by two to conform with the ISO requirement [7] that a Z-score of <2 shall indicate satisfactory performance, i.e,. fit for the purpose of the assay.

In the same way, data from assays for aluminium, copper, selenium and zinc were also evaluated.

Results

Quality specifications

Typical between-laboratory CVs for lead in blood and aluminium, copper, selenium and zinc in serum are shown in Table 2. The targets for satisfactory performance established and currently applied by the organisers of the EQAS are also given in Table 2.

No recommendations or proposals from professional organisations concerning analytical targets for aluminium, copper, selenium and zinc in serum could be found. The US National Committee for Clinical Laboratory Standards and the US Centers for Disease Control and Prevention recommend for lead in blood that the specification for internal quality control limits should be ±20 μg/L or ±10%, whichever is greater [10, 11].

Intra-individual variation of lead in blood was investigated by Delves et al. [12] who reported on the temporal stability of blood lead concentrations of 21 healthy adults (14 men and 7 women) exposed only to environmental lead. A serial collection of 9–17 blood samples was obtained over 7–11 months. The average blood lead concen-

Table 2 Targets for satisfactory performance for Pb in blood and Al, Cu, Se, Zn in serum used by organisers of external quality assessment schemes, and typical between-laboratory coefficients of variation calculated from results of participants

	Blood Pb	Serum			
		Al	Cu	Se	Zn
Targets					
Concentration (μmol/L)	2.0	3.7	20.0	2.0	20.0
Allowable deviation (%)	20–30	10–23	7.5–10.0	10.0	7.5–10.0
Between-laboratory CV (%)	10–16	5–20	7.2–10.0	9.4–20.5	6.5–19.1

Table 3 Intra- and inter-individual biological variation of Cu, Se and Zn in serum/plasma (CV%)

Author	Study	Cu	Se	Zn
Gallagher et al. 1989 [14]	Intra-individual			
	day	2.88	–	4.36
	week	4.46	–	6.11
	month	4.93	–	9.00
Gonzalez-Revalderia et al. 1990 [15]	Intra-individual	5.6	–	9.3
Ricos et al. 1999 [16]	Intra-individual			
	serum	4.9	12.0	11.0
	plasma	8.0	12.0	9.3
Lacher et al. 2005 [17]	Intra-individual	–	5.1	–
Sabban 2005 [18]	Intra-individual			
	day	2.3	4.4	9.4
	week	3.4	3.6	7.5
Gallagher et al. 1989 [14]	Inter-individual			
	day	12.18	–	8.07
	week	13.13	–	3.87
	month	13.83	–	4.07
Gonzalez-Revalderia et al. 1990 [15]	Inter-individual	13.6	–	9.4
Ricos et al. 1999 [16]	Inter-individual			
	serum	13.6	12.0	14.0
	plasma	19.0	14.0	9.4
Lacher et al. 2005 [17]	Inter-individual	–	13.2	–
Sabaan 2005 [18]	Inter-individual			
	day	15.9	13.9	10.7
	week	15.6	13.0	8.2

tration (in 1982) was 0.58 ± 0.11 μmol/L (18.9%) and the intra-individual variation, ranged from 1.4 to 9.1%, with an average value of 4.5%. In a similar study where serum aluminium concentrations were measured in subjects with normal renal function at bi-weekly intervals for 30 weeks, the intra-individual variation was <0.19 μmol/L [13]. Intra- and inter-individual variation for copper, selenium and zinc in serum, have been investigated by Gallagher et al. [14], Gonzalez-Revalderia et al. [15], Ricos et al. [16], Lacher et al. [17] and Sabban [18]. The copper, selenium and zinc data are summarised in Table 3. It can be seen that conflicting results were reported for some situations and the further investigation is required. Nevertheless, the more consistent results (which were obtained using the most recent technology), together with the data for lead and aluminium, were applied to the Fraser formulae.

The resulting TEa outcomes for minimal, desirable and optimal performance targets for copper, selenium and zinc are shown in Table 4. Taking into consideration the analytical data summarised in Table 2, it is apparent that the optimal and desirable values are achievable by very few of the participants and it is concluded that, at the moment, only the minimal TEa values can be employed as quality specifications. To simplify the implementation of these specifications and their understanding by participants, the Network members agreed that the value for zinc (i.e. ± 15%) be applied to all three of these measurements. It is important to note that the data evaluated refer to concentrations within the reference ranges for healthy subjects and that the specifications may need to be modified for lower levels. Table 5 summarises the proposed quality specifications for these elements and those for lead in blood and aluminium in serum, which are based on the desirable TEa values. Full details of the calculations, as applied to lead and aluminium, were given in earlier work [3].

Table 4 Minimal, desirable and optimal targets for total allowable error (TEa%) for measurements of Cu, Se and Zn in serum

	Cu	Se	Zn
Minimal	12	12	15
Desirable	8	8	10
Optimal	4	4	5

Table 5 Proposed quality specifications for Pb in blood and Al, Cu, Se, Zn in serum

Assay	Quality specification
Lead in blood	± 40 μg/L or ± 10%, whichever is the greater
Aluminium in serum	± 5 μg/L or ± 20%, whichever is the greater
Copper in serum	± 15% (at concentrations greater than 10 μmol/L)
Selenium in serum	± 15% (at concentrations greater than 0.7 μmol/L)
Zinc in serum	± 15% (at concentrations greater than 10 μmol/L)

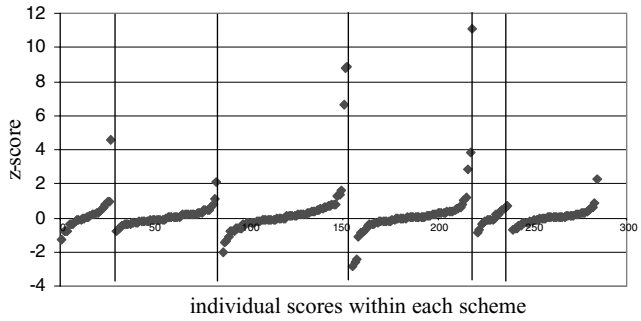

Fig. 1 Z-scores for results from six separate schemes for lead in blood. All samples had concentrations close to 100 μg/L (0.5 μmol/L)

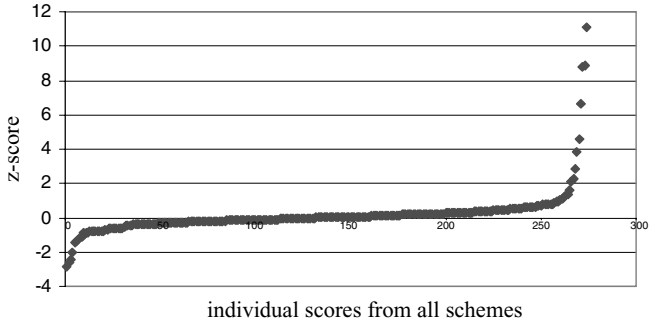

Fig. 2 An amalgamation of the Z-scores for results from six separate schemes for lead in blood. All samples had concentrations close to 100 μg/L (0.5 μmol/L)

Assessment of laboratory performance

The Z-scores for each of the blood lead results from the six schemes are shown in Fig. 1 and are combined in Fig. 2. Similar graphs can be prepared using the data for the other elements. It can be seen that there are small numbers of extreme outliers, which are likely to be due to non-analytical errors (blunders) e.g. transcription mistakes, analysis of the wrong specimen, or reporting in the wrong units. These outliers are not confined to one or two schemes, but occur in most. The percentage of participants gaining Z-scores within the range $\leq \pm 2$, i.e. gave results that can be defined as fit-for-purpose were: blood lead at 100 μg/L; 95%. Serum aluminium at 5.55 μmol/L; 73%. Serum copper at 16.5 μmol/L; 90%. Serum selenium at 1.52 μmol/L; 82%. Serum zinc at 20.6 μmol/L; 79%.

Discussion

The reasons for different scoring systems among similar schemes are complex. Some are prescribed within national legislation and the organisers have no authority to introduce any amendments. Others were established many years ago and are familiar to the participants and organisers, and are therefore reluctant to introduce change. In some countries, a particular scoring procedure is employed in large general clinical schemes and the organisers of smaller schemes are constrained to adopt the same scoring procedure in the interests of harmonisation and, again, for the convenience for participants.

The organisers of the schemes represented by the network of EQAS in OELM have shown that it is possible to agree upon a procedure for assessing performance that is consistent with the recommendations of ISO Guide 43-1 [7] and give Z-scores that provide equivalence of performance assessment across schemes. By using quality specifications based on the hierarchy outlined by Kenny et al. [2], with biological variability to define an agreed upon denominator, 'σ_p', performance that is satisfactory (i.e. $z \leq 2$) is shown to be fit for the intended purpose.

With the proposal presented here, organisers may:

- continue with their existing procedure, but have the new scores available to relevant authorities and/or accreditation bodies
- continue with their existing procedure, but use the new scores in collaborative work with other scheme organisers
- continue with their existing procedure, but also inform participants that a new score is available and indicate to them the advantages of assessment in this way. Participants may then choose to receive the revised performance assessment
- plan to discontinue their existing procedure and introduce the new scores

Success of this proposed harmonised approach requires that all scheme organisers are confident of the quality of the specimens that are distributed to participants. If there are concerns about the homogeneity, stability or possible contamination of test materials it will not be possible to be assured that any apparent differences between performances of laboratories in schemes are real.

The practical application of these quality specifications for lead in blood and for aluminium, copper, selenium and zinc in serum remains to be fully implemented. When in place, assessment of performance in this way in EQAS will immediately demonstrate whether results from participants can be regarded as fit for the purpose for which they are undertaken. One scheme already uses the proposed specifications for lead and aluminium to judge the performance of participants while those used for copper, selenium and zinc in serum are close to those that are proposed here [19]. Further work to evaluate the impact of these proposed quality specifications on the assessment of performance of participants in other schemes is now being undertaken so that organisers can determine whether to revise their scoring protocols in the near future. At the same time, work is being carried out to define suitable specifications for low concentrations of copper, selenium and zinc in serum.

Similar collaboration among the EQAS/PT scheme organisers of other sectors is recommended to provide the dual objectives of using a Z-score that clearly demonstrates when a laboratory is achieving results that are fit for purpose and allowing for comparison of performance from one scheme to another.

Acknowledgements This work was supported in part by the Commission of the European Union under Contract No. SMT4-CT98-7522

References

1. Fraser CG (1999) Scand J Clin Lab Invest 59:487–490
2. Kenny D, Fraser CG, Hyltoft Petersen, Kallner A (1999) Scand J Clin Lab Invest 59:585
3. Taylor A, Angerer J, Claeys F, Kristiansen J, Mazarrasa O, Menditto A, Patriarca M, Pineau A, Schoeters I, Sykes C, Valkonen S, Weykamp C (2002) Clin Chem 48:2000–2007
4. Christensen JM, Olsen E (1996) Ann Ist Super Sanità 32:285–294
5. Van Blerk M, Libeer J-C (2003) EQANews 14:62–66
6. Boley N (2005) Comparability of PT schemes – what did we learn from COEPT? Eurachem 5th Workshop Proficiency Testing in Analytical Chemistry, Microbiology and Laboratory Medicine, Portoroz, Slovenia. Accred Qual Assur 2005 (this issue)
7. International Organization for Standardization (1997) ISO Guide 43: Proficiency testing by interlaboratory comparisons, Part 1: Development and operation of proficiency testing schemes. Geneva, Switzerland
8. International Organization for Standardization (2005) ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva, Switzerland
9. Thompson M, Ellison S, Wood R (2006) The international harmonized protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure and Applied Chem (in press)
10. National Committee for Clinical Laboratory Standards (2001) Analytical Procedures for the Determination of Lead in Blood and Urine; Approved Guideline. NCCLS document C40-A, National Committee for Clinical Laboratory Standards, Wayne, PA, USA
11. Centers for Disease Control and Prevention (1997) Screening Young Children for Lead Poisoning: Guidance for State and Local Public Health Officials, Centers for Disease Control and Prevention, Atlanta, GA. Appendix C.1
12. Delves HT, Sherlock JC, Quinn MJ (1984) Hum Toxicol 3:279–288
13. Nordal KP, Dahl E, Thomassen Y, Brodwall EK, Halse J (1988) Pharmacol Toxicol 62:80–83
14. Gallagher SK, Johnson LK, Milne DB (1989) Clin Chem 35:369–373
15. Gonzales-Revalderia J, Garcia-Bermejo S, Menchen-Herreros A, Fernandez-Rodriguez E (1990) Clin Chem 36:2140–2141
16. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M (1999) Scand J Clin Lab Invest 59:491–500
17. Lacher D, Hughes J, Carroll M (2005) Clin Chem 51:450–452
18. Sabban A (2005) Quality specifications for copper, selenium and zinc in serum. MSc Thesis, University of Surrey, Guildford, Surrey, UK
19. www.wadsworth.org/testing/lead/ptprogram.htm