

Opinion Paper

Graham R.D. Jones*, Stephanie Albaredo, Dagmar Kessler, Finlay MacKenzie, Joy Mammen, Morten Pedersen, Anne Stavelin, Marc Thelen, Annette Thomas, Patrick J. Twomey, Emma Ventura and Mauro Panteghini, for the EFLM Task Finish Group – Analytical Performance Specifications for EQAS (TFG-APSEQA)

Analytical performance specifications for external quality assessment – definitions and descriptions

DOI 10.1515/cclm-2017-0151

Received February 21, 2017; accepted April 18, 2017; previously published online May 23, 2017

Abstract: External Quality Assurance (EQA) is vital to ensure acceptable analytical quality in medical laboratories. A key component of an EQA scheme is an analytical performance specification (APS) for each measurand that a laboratory can use to assess the extent of deviation of the obtained results from the target value. A consensus conference held in Milan in 2014 has proposed three models to set APS and these can be applied to setting APS for EQA. A goal arising from this conference is the harmonisation of EQA APS between different schemes to deliver consistent quality messages to laboratories irrespective of location and the choice of EQA provider. At this time there are wide differences in the APS used in different EQA schemes for the same measurands. Contributing factors to this variation are that the APS in different schemes are established using different criteria, applied to different types of data (e.g. single data points, multiple data points), used for different goals (e.g. improvement of analytical quality; licensing), and with the aim of eliciting different responses from participants. This paper provides recommendations from the European Federation of

Laboratory Medicine (EFLM) Task and Finish Group on Performance Specifications for External Quality Assurance Schemes (TFG-APSEQA) and on clear terminology for EQA APS. The recommended terminology covers six elements required to understand APS: 1) a statement on the EQA material matrix and its commutability; 2) the method used to assign the target value; 3) the data set to which APS are applied; 4) the applicable analytical property being assessed (i.e. total error, bias, imprecision, uncertainty); 5) the rationale for the selection of the APS; and 6) the type of the Milan model(s) used to set the APS. The terminology is required for EQA participants and other interested parties to understand the meaning of meeting or not meeting APS.

Keywords: analytical performance specifications; external quality assurance; international harmonisation; proficiency testing.

Background

External Quality Assurance (EQA) is one of many processes used to ensure the analytical quality of laboratory

***Corresponding author: Graham R.D. Jones,** Department of Chemical Pathology, St Vincent's Hospital, Sydney – SydPath, Victoria St, Darlinghurst 2010, Australia, Phone: +61 2 83829160, E-mail: Graham.Jones@svha.org.au; University of NSW, Sydney, Australia; and RCPAQAP, Sydney, Australia

Stephanie Albaredo: Centre Toulousain pour le Contrôle de qualité en Biologie clinique (CTCB), Toulouse, France

Dagmar Kessler: Quality Control Centre Switzerland (CSCQ), Geneva, Switzerland

Finlay MacKenzie: Birmingham Quality, UK NEQAS, University Hospitals Birmingham NHS Trust, Birmingham, UK

Joy Mammen: Department of Transfusion Medicine, Christian Medical College Vellore, Vellore, Tamilnadu, India

Morten Pedersen: Danish Institute for External Quality Assurance in Laboratory Medicine (DEKS), Copenhagen University Hospital, Glostrup, Denmark

Anne Stavelin: The Norwegian Quality Improvement of Laboratory Examinations (Noklus), Haralds plass Deaconess Hospital, Bergen, Norway

Marc Thelen: Laboratory for Clinical Chemistry, Amphia Hospital, Breda, The Netherlands; and SKML, Organization for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands

Annette Thomas: WEQAS, Cardiff and Vale University Health Board, Cardiff, UK

Patrick J. Twomey: Department of Chemical Pathology, St Vincent's University Hospital, Dublin, Ireland; and School of Medicine, University College Dublin, Ireland

Emma Ventura: Sociedad Española de Bioquímica Clínica y Patología Molecular (SEQC), Barcelona, Spain

Mauro Panteghini: Research Center for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy

measurements. The usual process is for an EQA provider to distribute proficiency test items or samples to participating laboratories (customers or participants), which perform a range of measurements and return the measurement results to the EQA organiser. The scheme organisers then provide a report to the participants that compares the submitted result(s) with a target value (assigned value). Analytical quality is assessed by considering the difference between the result(s) and the target value(s) assigned by the EQA scheme. EQA organisers provide analytical performance specifications (APS) that indicate whether the deviation from the target value achieved by the laboratory is acceptable. An APS is generally expressed as a number of units or a percentage deviation from a specified target, creating upper and lower acceptance limits. Other terms for APS that have been used include Performance Goals, Quality Specifications, Analytical Performance Goals, Quality Standards, Allowable Limits of Performance, Acceptability Limits and Quality Goals. The term Analytical Performance Specifications is preferred in line with the terminology used in the Milan conference 2014 [1].

Different EQA providers however use a wide range of APS for the same measurand [2–5]. A survey performed by the European Organisation for External Quality Assurance Providers in Laboratory Medicine (EQALM) in 2014 showed that the criteria used to set the APS also vary widely [6], which is presumably a strong contributor to the variation in APS. It was noted that there was a wide variation in the way APS were determined for EQA schemes, indicating that different information was being conveyed

to laboratories by the APS in different schemes [4]. Table 1 shows the variation in processes used to establish APS from some EQA schemes involving the authors.

There has been considerable work over the last 15 years in the general field of APS to improve their application in laboratory medicine. The concepts were initially codified in the so-called Stockholm hierarchy that outlined a structured approach to setting APS [7]. There have been a number of demonstrations of the application of these principles to EQA [5, 8, 9]. In 2014, these criteria were revisited in Milan where a set of three models were proposed for establishing APS [1]. As a follow-on from the Milan meeting, the EFLM convened a number of task and finish groups (TFG) to address issues arising from the meeting [10]. The present paper is a product of the TFG on APS for EQA.

Since EQA schemes are designed differently, the EQA APS are based on different criteria, often aiming to convey different information about assay and/or participant performance. These variations increase the complexity of making a comparison between the APS provided by different EQA organisers. There are also different designs of EQA schemes that are also relevant to interpretation of results [11].

The aim of this document is to define, describe and facilitate communication of the essential components of EQA APS in the field of quantitative laboratory medicine testing. It is our intention that this information will assist EQA organisers in establishing APS and then providing descriptions of the APS to participants and other stakeholders. This also provides a background for the development

Table 1: Examples of current variation in models used to assign analytical performance specifications (APS) to External Quality Assurance (EQA) schemes.

| EQA Program | Models |
|----------------------|---|
| CSCQ Switzerland | Governmental regulations (combination of BV and state of the art) and Combination of limits given by scientific societies and Z-score |
| CTCB France | Z-score/state of the art/limits given by scientific societies or other/limits based on clinical impact |
| DEKS Denmark | Combination of BV, state of the art and expert opinion |
| NOKLUS Norway | Fixed percentage limits and based on a combination of BV, state of the art and expert opinion |
| RCPAQAP Australia | Combination of BV and state of the art |
| SEHH Spain | Statistical/state of the art/BV |
| SEQC Spain | Combination of BV and statistical results |
| SKML The Netherlands | Combination of BV and state of the art |
| WEQAS UK | Combination of BV and state of the art |
| CMCEQAS | Combination of state of the art and statistical considerations |

CSCQ, Suisse de Contrôle de Qualité; CTCB, Centre Toulousain pour le Contrôle de qualité en Biologie Clinique; DEKS, Danish Institute of External Quality Assurance for Laboratories in Health Care; NOKLUS, Norwegian Quality Improvement of laboratory examinations; RCPAQAP, Quality assurance Program of the Royal College of Pathologists of Australasia; SEHH, Spanish Society of haematology and haemotherapy; SEQC, Spanish Society of Clinical Biochemistry and Molecular Pathology; SKML, Dutch Foundation for Quality Assessment in Medical Laboratory; WEQAS, Welsh EQA provider; CMCEQAS, Christian Medical College External Quality Assurance Scheme; BV, biological variation.

of APS, which should be set by laboratory professionals using a model for each measurand selected from the Milan consensus statement [1] using concepts that have been defined by another EFLM TFG and are now available [12].

Results for qualitative, semi-quantitative or morphological examination are not dealt with in this document. The recommendations do not apply to EQA schemes scores based on combined results from multiple measurands or non-analytical aspects, such as participation rate, delayed responses, number of amendments, and so forth.

The target audience for this document includes EQA organisers, EQA participants and potential participants, accreditation bodies, competent authorities, IVD manufacturers and laboratory professional organisations and has been prepared to be compliant with the relevant ISO documents, ISO17043:2010 [13] and ISO13528:2015 [14].

Recommendations

The recommendations below are relevant to any setting supported by EQA, including laboratory, point-of-care testing and physician office. It is recognised that multiple APS may be applied to the same result(s) for the same measurand. In this case, the information must be supplied for each APS.

The EFLM TFG-APSEQA has identified six important elements to facilitate the description and communication of APS.

A. Elements of description of the APS

1. The nature of the EQA material

The EQA material matrix and its commutability should be specified. This is because the interpretation of differences between results in a scheme is dependent on the nature of the material. Examples of such description may be summarised as:

1. Material known to be commutable (information of the process used to establish commutability should be available)
2. Material likely to be commutable (e.g. fresh serum without additions, however, commutability not assessed experimentally)
3. Material known not to be commutable
4. Specific limitations (e.g. if a material is known to be generally commutable but non-commutable for one or more methods or one or more measurands)
5. Commutability not assessed

2. The procedure used to establish the assigned value

The organiser must state the procedure(s) followed to establish the assigned value. When comparing the difference between a result and the assigned value, it is necessary to be aware of any limitations and uncertainty due to the nature of the process. Examples of these procedures include:

1. Measurement of the EQA material with a reference measurement procedure by a reference laboratory
2. Comparison with a certified reference material
3. Formulation (weighed-in values), e.g. for exogenous measurands such as therapeutic drugs
4. Derived from the submitted results from the scheme by a described statistical process. Such assigned values may be for all laboratories in the survey or for specific subgroups, e.g. based on measuring system, reagent manufacturer, instrument or analytical method. An example would be the all laboratory trimmed mean
5. Derived from the submitted results of an expert panel from the scheme by a described statistical process

3. The data set for application of APS

EQA providers must provide information about the data set to which the APS are applied.

The following are examples of descriptions of the data set to which the APS can be applied:

1. To a result from single measurements on a single specimen
2. To n separate results from single measurements of multiple specimens
3. To the average of n multiple measurements on a single specimen (e.g. if samples are measured in duplicate and the average submitted)
4. To results from specified method groups

The number of data points included in these calculations will affect the uncertainty of the calculations.

4. The applicable analytical quality being assessed

The organiser should state which aspect(s) of the analytical quality are being evaluated

The aspects of analytical quality usually assessed by EQA are total error, bias and imprecision. Bias and imprecision can only be determined by calculation based on a number of measurements. Assessment of

performance based on a single result is by necessity using a total error APS; as bias and imprecision cannot be separated. The aspect of quality for the APS may be described as follows:

1. Total error (includes bias and imprecision, as applied to single result or calculated from multiple results)
2. Bias (may be expressed as absolute or relative bias of one or more samples as a single value or as a mathematical equation reflecting the relation between concentration level and the measurement[s])
3. Imprecision (expressed as absolute or relative to the concentration level)

The mathematical approach that is used to calculate total error, bias and imprecision should be documented.

5. The rationale for the selection of the APS

The organiser must state the purpose for which the APS are intended. The rationale behind the APS affects the way an EQA organiser establishes the limits and is related to the expected or required response by participants to a failure for a result to meet the APS. Examples may include one or more of:

1. Passable (everyone should theoretically pass; there may still be clinical benefit from better performance. Regulatory requirements or governmental regulations may favour this philosophy)
2. Satisfactory (good performing laboratories should pass; this philosophy is oriented to maintain current performance)
3. Favourable (no clinical benefit of further improvement)
4. Aspirational (aim to improve performance, educational)

6. Type of model for establishing the APS

The EQA organiser must state the model used to establish the APS. It is recommended that one of the models from the Milan conference is used [1] although it is also recognised that data from different models may be used to establish a final APS, e.g. state of the art may be used to determine which category within biological variation is selected (optimal, desirable, minimal). These can be described as:

1. Outcome-based (Milan model 1a)
2. Based on clinical decision applications (Milan model 1b)
3. Derived from biological variation (Milan model 2)

4. State of the art, defined as the highest level of analytical performance technically achievable in that moment (Milan model 3)

B. Communication with EQA participants/ stakeholders

It is recommended that EQA organisers provide a summary of their APS as well as a detailed description of the elements listed above in a standardised format. This should be provided in language(s) for the intended participants as well as in English and be openly available to other interested parties. Since different measurands in the same scheme, and the same measurands in different schemes, may have different APS descriptions at least in some perspectives, APS descriptions should be made available for each measurand in every scheme.

Discussion and conclusions

The response of laboratories to EQA reports is influenced by the APS provided by the EQA organiser. The same result(s) may be either accepted or further investigated depending on the APS in place. Given this importance, the processes used to set APS in EQA and the communication of the use and meaning of EQA APS should be clarified by the implementation of the structured approach and terminology recommended in this paper. The elements listed above to describe APS are all considered necessary to make a fully informed assessment of the analytical performance of a laboratory based assessment of EQA results using supplied APS.

The nature of the material must be known to ensure it is appropriate for the comparison being made (element 1). If a comparison is made with results derived from different measurement techniques, including reference methods, then a knowledge of the commutability is required for correct interpretation. The details of the process for value assignment of the target is also required (element 2). An assigned value with a different traceability chain to the laboratory's method, or with a large uncertainty, will influence the interpretation of the result. Supplying information about the uncertainty of the assigned value may also be of use to program participants. A valid comparison with a higher-order reference method using a commutable sample requires the information described in elements 1 and 2.

The data to which the APS are applied must be clearly defined (element 3). Within a program report, averages of

multiple results or multiple measurements, or other ways of combining results may be handled seamlessly. However when interpreting the factors which have produced an abnormal result a clear understanding of the components is required. Additionally awareness of the structure of a program, i.e. the number of samples in a survey and the frequency of the surveys, allows each result to be interpreted in the context of the other available results.

It is also important to understand the analytical quality to which an APS is applied (element 4). Information derived from measurements of multiple samples permits the assessment of laboratory bias or imprecision, as opposed to single results where total error is assessed. The use of total error limits to assess bias, for example, could lead to a misinterpretation of the assay quality to be better than it actually is because the APS would be too wide.

A knowledge of the rationale behind setting the APS is also required for correct interpretation of the EQA result (element 5). A result inside a wider limit (e.g. regulatory) may pass this criteria, but not be optimal for patient management. Alternatively a failure to meet tighter limits (e.g. aspirational) may be due to limitations in the available methods rather than individual laboratory performance.

Finally, knowledge of the model used to establish the APS can affect interpretation (element 6). As well as the three models described in the Milan consensus, grading (e.g. optimal, desirable and minimal) is highly recommended with clear definitions of the grades required. If

a method meets the optimal level relative to biological variation, or meets a defined clinical need, then spending time considering further improvement is unnecessary. However, if the APS are based on state of the art, or minimal standards for biological variation, then further improvement may be of benefit to patients. It is recognised that currently EQA organisers often use models not included in the Milan consensus (Table 1). While these should be noted as such, organisers are encouraged to base decisions on APS on the Milan models, recognising that there is progress in the field of assigning measurands to the various models.

As stated in the preceding paragraphs all the elements listed in this paper are required for a considered interpretation of an EQA result against the supplied APS. An additional reason for a detailed description of required information is to allow comparison of APS between schemes from different providers. If APS from one provider are known to be based on state of the art and from another are set based in desirable biological variation, then differences can be explained and customers can be aware of the reason for apparent different performance in different schemes.

An example of summary supporting information is shown in Table 2 based on the RCPAQAP from Australia. The table provides the required information and demonstrates that APS may differ between measurands in the same EQA scheme. The table includes a reference to a detailed description of the process used to establish the APS.

Table 2: Example of summary description of analytical performance specifications (APS) based on the RCPAQAP General Serum Chemistry External Quality Assurance (EQA) Scheme.

1. The EQA material is not validated as commutable
2. The overall target-setting method for each measurand is shown below. In addition, method, instrument, reagent manufacturer-based consensus targets are provided based on returned results
3. The APS are to be applied to each individual measurement result
4. The APS are applied for assessment of total error (i.e. the effects of imprecision and bias combined)
5. The rationale for the APS is ‘Aspirational’ (to improve performance) where this is required. The response of the laboratory to ‘out of range’ results should be to review performance and seek improvement
6. The APS are established based on biological variation and state of the art (levels 2 and 3 from Milan conference). The components of biological variation and the level (optimal, desirable, or minimal) are shown below

Further details on the RCPAQAP process used to establish these APS are available [9, 15]

| Measurand | Assignment of target | Analytical performance specifications | Employed component(s) of biological variation | Quality level |
|-----------------|---|---|---|---------------|
| S/P-ALT | IFCC reference procedure in a JCTLM-listed reference laboratory | ±5 U/L up to 40 U/L; ±12% >40 U/L | Within-individual (imprecision) | Optimal |
| S/P-Bicarbonate | Selected well-controlled commercial measuring system by an ISO 15189 accredited clinical laboratories | ±2.0 mmol/L up to 20.0 mmol/L; ±10% >20.0 mmol/L | Within- and between-individual (total error) | Minimal |
| S-Transferrin | Median of laboratories participating in EQA | ±0.20 g/L up to 2.50 g/L; ±8% >2.50 g/L | Within- and between-individual (total error) | Minimal |

In setting APS, EQA providers should take into account the required response from participants who fail to meet specifications. This aspect reflects the intention of the information conveyed by passing/failing relative to an APS. For the same reason, EQA providers should clearly indicate when the APS elements differ between schemes or between measurands so that participants are aware of any such differences. For example in Table 2, different analytes in the same scheme use APS based on total error as well as precision criteria, with some at the optimal level and some minimal.

The ISO standard for clinical laboratories, ISO 15189, [16] requires that laboratories validate or verify the performance of a measurement procedure for the ‘intended use’. Since a participant may apply a test for a different use than was envisaged by the EQA provider, the APS of a particular scheme may be not applicable to their situation. For instance, if a laboratory applies a certain glucose test only to separate hypoglycaemic from hyperglycaemic comatose patients in the intensive care unit from its hospital, wider APS can be applicable than for other applications of glucose testing (e.g. diabetes diagnosis). As EQA organisers cannot have APS for every possible intended use of a test, laboratories are recommended to document their own required response to results if their use of the assay differs from generally expected use.

The final goal of laboratory medicine is enabling high quality medical decision making. One aspect of this is understanding the effect that the quality of laboratory data has on the manner in which it can be used in patient care. EQA data can be employed to inform how laboratory measurement results are clinically suitable and should be interpreted correctly. For example, data for a measurand from laboratories with a small variation relative to the within-subject biological variation of the measurand can be safely used to monitor a patient. Data from laboratories with a large variation relative to the within-subject biological variation will conversely increase the noise seen by the clinician such that larger changes in results might be required to be certain of a significant change in a patient’s clinical status. Similarly, results from laboratories with a clinically insignificant bias may be able to share a common reference interval or decision limit. However, laboratories with a clinically significant bias should not use common reference intervals or clinical decision limits.

In conclusion, we consider it indispensable that EQA schemes advise participants and all interested stakeholders about the nature of their provided APS in sufficient detail to allow informed decisions about the meaning of the results, as well as to allow valid comparison of APS from different EQA schemes.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones GRD, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st strategic conference of the European federation of clinical chemistry and laboratory medicine. *Clin Chem Lab Med* 2015;53:833–5.
2. Ricos C, Baadenhuijsen H, Libeer CJ, Petersen PH, Stockl D, Thienpont L, et al. External quality assessment: currently used criteria for evaluating performance in European countries, and criteria for future harmonization. *Eur J Clin Chem Clin Biochem* 1996;34:159–65.
3. Stavelin A, Meijer P, Kitchen D, Sandberg S. External quality assessment of point-of-care international normalized ratio (INR) testing in Europe. *Clin Chem Lab Med* 2012;50:81–8.
4. Jones GR. Analytical performance specifications for EQA schemes – need for harmonisation. *Clin Chem Lab Med* 2015;53:919–24.
5. Thelen M, Jansen R, Weykamp C, Steigstra, H, Meijer, Cobbaert C. Expressing analytical performance from multi sample evaluation in laboratory EQA. *Clin Chem Lab Med* 2017. DOI 10.1515/cclm-2016-0970. [Epub ahead of print].
6. Albe X. How is poor performance defined among EQA organisations. EQALM meeting Toulouse 2014. Available at http://www.eqalm.org/site//2014/1-2_X_Albe_Poor_Performance.pdf (accessed 7th Feb 2017).
7. Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Strategies to set global analytical quality specifications in laboratory medicine – consensus agreement. *Scand J Clin Lab Invest* 1999;59:585.
8. Sciacovelli L, Zardo L, Secchiero S, Plebani M. Quality specifications in EQA schemes: from theory to practice. *Clin Chim Acta* 2004;346:87–97.
9. Jones GRD, Sikaris K, Gill J. ‘Allowable limits of performance’ for External Quality Assurance programs – an approach to application of the Stockholm criteria by the RCPA quality assurance programs. *Clin Biochem Rev* 2012;33:133–9.
10. Panteghini M, Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. *Clin Chem Lab Med* 2015;53:829–32.
11. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* 2011;57:1670–80.
12. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined

- in the 1st EFLM Strategic Conference. Clin Chem Lab Med. 2017;55:189–94.
13. ISO 17043:2010. Conformity assessment – General requirements for proficiency testing. ISO/CASCO – Committee on conformity assessment. 2010.
 14. ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparison. ISO TC 69. 2015.
 15. Jones GR. Common performance specifications in EQA – is it possible? EQALM meeting, Bergen, 2015. Available at http://www.eqalm.org/site//2015/3-Performance%20specifications%20in%20EQA_Jones.pdf (accessed 9th April 2017).
 16. ISO 15189:2012. Medical laboratories – Particular requirements for quality and competence. ISO TC 212. 2012.