Opinion Paper

Tony Badrick* and Peter Graham

Can a combination of average of normals and “real time” External Quality Assurance replace Internal Quality Control?

<https://doi.org/10.1515/cclm-2017-0115>

Received February 8, 2017; accepted August 9, 2017; previously published online September 13, 2017

Abstract: Internal Quality Control and External Quality Assurance are separate but related processes that have developed independently in laboratory medicine over many years. They have different sample frequencies, statistical interpretations and immediacy. Both processes have evolved absorbing new understandings of the concept of laboratory error, sample material matrix and assay capability. However, we do not believe at the coalface that either process has led to much improvement in patient outcomes recently. It is the increasing reliability and automation of analytical platforms along with improved stability of reagents that has reduced systematic and random error, which in turn has minimised the risk of running less frequent IQC. We suggest that it is time to rethink the role of both these processes and unite them into a single approach using an Average of Normals model supported by more frequent External Quality Assurance samples. This new paradigm may lead to less confusion for laboratory staff and quicker responses to and identification of out of control situations.

Keywords: frequency of quality control samples; quality assurance; real-time EQA.

Introduction

Quality Management Systems such as ISO 15189 dictate that laboratories shall use Internal Quality Control (IQC) procedures and participate in External Quality Assurance (EQA) to ensure that the results of the tests they report are

fit for purpose [1]. IQC and EQA are mutually dependent but often unrelated activities. IQC has a well-defined statistical basis for the rules to be used; however, the frequency of EQA challenges and determination of allowable limits varies widely across the available programs. Even the aims of different programs are different depending on whether it is a Proficiency Testing program or an EQA Program.

Improvements in the stability of analytical methods and the availability of more targeted routines to determine Average of Normals (AoN) has resulted in longer timeframes between IQC checks [2–4]. Most laboratories now run IQC samples only once or twice a day, although this can lead to problems if there are catastrophic failures and samples need to be rerun. Guidelines are available to implement and monitor IQC [5], yet there is evidence that laboratories still do not perform IQC effectively [6, 7]. Likewise, a number of EQA providers have reduced the timeframe between EQA challenges with more samples and improved (online) reporting.

The adoption of AoN requires laboratories to reassess the role of IQC and EQA, which will in turn change the way EQA is delivered. Using a combination of AoN and real-time EQA will change the way laboratories manage their quality.

AoN methods offer almost real-time monitoring of analytical bias, but this is only one dimension of analytical performance. IQC and EQA/PT schemes assess analytical bias and imprecision compared with peer method groups. Although preanalytical and postanalytical errors can be very difficult to identify and control for, analytical variability (both imprecision and inaccuracy) can be monitored with IQC programs and EQA schemes, yet many laboratories run IQC and EQA almost independently and fail to link performance between the two. We will argue that the time has come to integrate EQA and AoN to replace conventional IQC in many routine diagnostic laboratory situations.

First, we need to define the type of EQA scheme we need in place to achieve this integration.

The requirements of the necessary EQA program we seek have been described by Sturgeon [8] and include

*Corresponding author: Tony Badrick, RCPA Quality Assurance Programs, St Leonards, Sydney 2065, Australia, E-mail: tony.badrick@rcpaqap.com.au

Peter Graham: RCPA Quality Assurance Programs, St Leonards, Sydney, Australia

Table 1: A comparison of the frequency of challenges in some EQA programs.

Provider	Program	Surveys per year	Samples per survey	Challenges per year	Frequency of analysis and reporting
BioRad	Chemistry and Immunoassay	24	1	24	2 weeks
RIQAS	Chemistry	26	1	26	2 weeks
UKNEQAS	Chemistry	24	3	72	2 weeks
UKNEQAS	HbA _{1c}	12	3	36	4 weeks
UKNEQAS	Specific Proteins	12	2	24	4 weeks
CAP	General Chemistry	3	5	15	4 months
CAP	Endocrinology	3	5	15	4 months
CAP	Lipids	2	3	6	6 months
CAP	*Calibration verification/Linearity sets	2	7	14	6 months
THISTLE programs (CEQAL)	Routine Chemistry	3	5	15	4 months
RCPAQAP	General Serum Chemistry	24	2	48	2 weeks
RCPAQAP	Condensed Chemistry Program	12	2	24	2 weeks

BioRad, BioRad Laboratories; RIQAS, Randox International Quality Assessment Scheme; UKNEQAS, United Kingdom National Quality Assessment Service; CAP, College of American Pathologists; CEQAL, Canadian External Quality Laboratory; RCPAQAP, Royal College of Pathologists of Australasia Quality Assurance Programs.

assessment of overall performance, assessment of individual performance, assessment of method robustness to clinically relevant interference, communication with participating laboratories and audit of wider aspects of analytical performance and educational activities.

Requirements of EQA schemes

EQA programs vary in a range of different factors in their design (Table 1). These include the following:

- The number of samples over a period of time
- The number of samples tested at the same time (a testing episode)
- The range of concentrations assessed
- The number of samples across the concentration range
- The number of samples combined for statistical analysis

These factors are interdependent. For example, if the total number of samples in a time period remains the same, more measurements at one time lead to a greater time between measurement.

There is no current guideline or consensus on what the frequency of EQA challenges should be [9]. In fact, there is a wide range used in different programs (Table 1).

EQA versus IQC

The quality assurance system in clinical chemistry allows for the identification of errors and control actions to

correct them. The most important tool in this quality control is the use of control charts. Apart from providing real-time information on an out-of-control assay, the purpose of IQC is mainly to verify the stability of laboratory estimates with time, and therefore it is essentially aimed at detecting systematic drift/shift as would occur over time and requires recalibration. There is evidence that IQC has not provided a robust system of detection of such systematic errors, particularly with the advent of high-sensitivity immunoassays [10].

Although IQC and EQA schemes allow the laboratory to have a retrospective estimate of its performance in terms of both imprecision and inaccuracy and both assign targets where definitive or reference methods are available, IQCs more often provide method medians with a broad “acceptable range” quoted by manufacturers. Laboratories, instead, establish their own acceptable ranges based on previous experience and allowable performance limits.

Hence, the main feature that generally distinguishes EQA from IQC is accuracy. Determining a mean/target and acceptable limit on an unknown IQC material does not obviate the need to check for accuracy even when using a traceable calibrator. EQA providers strive to supply commutable samples with targets that are determined by reference methods using certified reference standards (where available). EQA samples can also be designed to test the “robustness” of methods and manufacturer claims by including samples with interferences (e.g. haemolysis, heterophilic antibodies [8]). Currently, the only feature that IQC provides that is not matched by EQA is “real time” alert to a potential issue (Table 2).

Table 2: Comparison of the essential differences between IQC and EQA programs.

Feature	EQA	IQC
Precision	✓	✓
Accuracy (of measurement)	✓	
Bias (systematic error)	✓	✓
Linearity	✓	
Limit of detection	✓	✓
“Robustness”, e.g. interferences	✓	
Peer comparison	✓	✓
“Real time” alert		✓
Commutable material	To be tested	To be tested

There may also be problems with EQA samples because of their nature. When investigating a problem identified by an EQA sample, the following should be considered: clerical errors, methodological problems (carryover, reagent or calibrator variation), equipment problems, technical problems related to personnel and problems with the EQA material [11]. Potential matrix problems with the EQA (and IQC) material can have a significant impact on the way the two systems (IQC and EQA) can be integrated [12], although it has been shown that the commutability of samples is greater than may have been previously reported [13]. In order to detect a shift in patient results, laboratories require an EQA program with commutable samples and robust target setting procedures [14]. Whether the EQA system uses reference value-derived targets or medians of method groups or all groups clearly will have an impact on the interpretation of results. Other key characteristics of EQA programs are summarised by Miller et al. [11].

A significant problem with EQA and IQC samples is that they are often not treated the same as patient samples in the laboratory. EQA samples are readily identifiable and are often run when the analyser is deemed to be performing optimally; in contrast, AoN uses commutable patient samples.

Laboratories currently use IQC sera for other purposes than control of an assay. These include (a) Validation of a manufacturer’s assay, (b) Determination of Long term Measurement Uncertainty (MU) of an assay, (c) Performance of laboratories in a network where the same manufacturer’s platform/assay is in use and the same lot of IQC is used to troubleshoot assay problems in a laboratory and (d) Use of global IQC systems to troubleshoot a reagent lot. Validation of a manufacturers assay is best performed using a set of patient samples covering a broad range of concentration. The determination of long-term MU can be achieved with EQA samples and is currently offered in some programs [15]. Where there is a network of

instruments of the same or different manufacturers measuring the same measurand, control can still be achieved using a combination of AoN to detect local shifts in bias and imprecision [4, 10] and EQA to detect inaccuracy. If there is a concern about a specific reagent lot, the most effective tool is to use a range of patient samples to investigate lot-to-lot variation [16].

Integrating EQA and AoN

We will restrict ourselves in this discussion to high-volume routine clinical chemistry and immunoassay analysers, although haematology analysers would also be amenable to this approach, particularly since the concept of AoN is already a key component of IQC on these platforms. What do we require from an EQA program to be able to drop conventional IQC given that there is an effective AoN routine?

There are strategies available to support the frequency and number of IQC samples in a “run” based on the capability of the assay [14, 17], with perhaps the most difficult definition being that of a “run”. There is also the very real problem of controlling an analyser with many tests on board particularly if different IQC rules based on different assay capabilities are in use. However, generally, regardless of the IQC rule, the best way to detect error is simply to increase the frequency of quality check (IQC or EQA) – the power increases with number. The balance then becomes the statistical power desired versus what is practical and cost-effective. An effective AoN routine may solve these problems. The concept of AoN will invariably develop and become broader. In fact, there is currently an international program that monitors the daily medians of the results for 20 commonly measured analytes in serum or plasma from contributing laboratories around the world [18]. This program incorporated the concept of a population-based rather than a site-based AoN.

The EQA needs to be able to detect bias and imprecision. This can be achieved by a program that has reference method determined targets, a range of concentrations that spans the clinically significant spectrum and replicate measurements that allow an estimate of CV to be determined. When the EQA samples are linearly related and repeat in a random sequence within a cycle, this increases the statistical power of the challenge. The interpretation of the EQA result should also be based on biological variation criteria not statistical rules.

Although analysers are quite stable, there are high-risk situations when this stability can be affected, namely, after a calibration, reagent change, start up or equipment

maintenance. These are times when additional monitoring is necessary as a change in bias will not be detected by the AoN process soon enough to prevent reporting incorrect patient results. There are a few possible solutions: the use of a set of patient samples as a lot-to-lot check [16] or running stored EQA samples (e.g. frozen aliquots of leftover EQA material). Using the lot-to-lot strategy may be complex, but perhaps if the frequency of recalibration is sufficiently low, this may be the ideal solution to reduce bias shift.

The missing components of EQA compared with IQC are the low frequency of EQA samples and the slow turnaround of results. The latter short coming could be overcome by real-time EQA, where the results are immediately returned when they are submitted. This would impose some problems on the EQA program, but it is not insurmountable. The low frequency of EQA samples problem could be solved by weekly challenges. The weekly sample would be run on a fixed day or sooner if there was a problem with the assay. Perhaps more EQA samples could be run on poorer performing assays based on a sigma metric. EQA samples need to be treated, as far as possible, like a patient sample and analysed randomly at any time during the run.

We recognise the practical issues of “real time” EQA for analytes where a commutable reference method central value is not readily available. Traditionally, EQA providers use medians in these situations. We would propose the combination of a central value and a “running median”. The central value would be predetermined by a minimum of five “target” laboratories from each of the major method groups. Target laboratories would be recruited from consistently well-performing laboratories (e.g. top 20% in a given EQA program). The “running median” would guide on the “trueness” of the central value as data accumulates.

Another advantage of predetermined central values (be they reference method targets or medians) is that results that would otherwise be inappropriate for statistical evaluation would be flagged “real time”. Sample blunders, mix ups or clerical errors would be identified in real time and reported to the laboratory for confirmation.

The potential cost of additional samples/challenges (around 100 per year) is also a consideration. Generally, the marginal cost of the base material for high-volume clinical chemistry analytes (human serum and urine) is not the limiting factor; it is the spiking, processing, value assignment, packaging and handling costs that add to the expense. Technically, these would be similar for IQC material, so the overall cost differential (real-time EQA vs. IQC plus conventional EQA) may still be a cost effective alternative.

EQA or Competence Assessment

It is important to differentiate between EQA and Competence Assessments. Some qualitative or semiquantitative tests are very dependent on the operator such as fluorescence *in situ* hybridization (FISH), porphyrin testing and microscopy. With these tests, it is the competence of the operator that we need to assess. Hence, the “EQA” should be individual and aimed at a baseline performance of do no harm to the patient. The sample challenges should ensure that an operator will not miss a key abnormality and the frequency of testing should reflect the risk of misinterpretation of the test. Anyone who may perform the test, particularly those who may only perform the test in rare situations, must be in this competence program.

Conclusions

Quality Control strategies are continually changing to incorporate concepts such as six sigma and risk [19]; however, there is still confusion amongst laboratory practitioners [14]. Not all EQA programs are the same. To obtain the most from the program, the laboratory must understand the strengths and weaknesses of the program design and the material used to make the samples.

AoN methods currently are most suitable for high-volume assays on relatively stable population groups and are sensitive to changes in populations. Although there are some methods that can increase the effectiveness of these techniques, including the selection of a “normal” population or application of truncation limits to the data, that is, removing the patients with abnormal results [10], and use of simulated annealing algorithms [3], they do not completely exclude the possibility of an underlying patient population shift. However, there are solutions to this problem coming in the form of the next-generation electronic medical record. These systems promise to bring together different clinical databases that are traditionally organised into silos [20]. This opens up the possibility for the laboratory to match their laboratory trends with clinical information.

We trust that the questions we have raised in this paper will prompt further discussion, including the fundamental question of whether an appropriately designed EQA scheme in conjunction together with an appropriate AoN procedure can replace conventional IQC. To achieve this, EQA providers would need to implement real-time options and consider a higher frequency of sample challenges than most schemes currently run.

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. International Organisation for Standardisation ISO 15189. Medical laboratories – Particular requirements for quality and competence. http://www.iso.org/iso/catalogue_detail?Csnumber=56115. Accessed 3 Feb 2017.
2. Fleming JK, Katayev A. Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: for patients by patients. *Clin Biochem* 2015;48:508–13.
3. Ng D, Polito FA, Cervinski MA. Optimization of a moving averages program using a simulated annealing algorithm: the goal is to monitor the process not the patients. *Clin Chem* 2016;62:1361–71.
4. Liu J, Tan CH, Loh TP, Badrick T. Verification of out-of-control situations detected by “average of normal” approach. *Clin Biochem* 2016;49:1248–53.
5. Jones G, Calleja J, Chesher D, Parvin C, Yundt-Pacheo J, Mackay M, et al. Collective opinion paper on a 2013 AACB workshop of experts seeking harmonisation of approaches to setting a laboratory quality control policy. *Clin Biochem Rev* 2015;36:87–95.
6. Howanitz PJ, Tetrault GA, Steindel SJ. Clinical laboratory quality control: a costly process now out of control. *Clin Chim Acta* 1997;260:163–74.
7. Litten J. Applying sigma metrics to reduce outliers. *Clin Lab Med* 2017;37:177–86.
8. Sturgeon CM. External quality assessment of hormone determinations. *Best Pract Res Clin Endocrinol Metab* 2013;127:803–22.
9. Parvin C. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. *Clin Chem* 2008;54:2049–54.
10. Liu J, Tan CH, Badrick T, Loh TP. Moving sum of number of positive patient result as a quality control tool. *Clin Chem Lab Med* 2017;55:1709–14.
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. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. *Clin Chim Acta* 2003;327:25–37.
13. Rej R, Norton-Wenzel CS. Assessing analytical accuracy through proficiency testing: have the effects of matrix been overstated? *Clin Chem* 2015;61:433–4.
14. Westgard JO, Westgard SA. The quality of laboratory testing today: an assessment of σ metrics for analytic quality using performance data from proficiency testing surveys and the CLIA criteria for acceptable performance. *Am J Clin Path* 2015;125:343–54.
15. Matar G, Poggi B, Meley R, Bon C, Chardon L, Chikh K, et al. Uncertainty in measurement for 43 biochemistry, immunoassay, and hemostasis routine analytes evaluated by a method using only external quality assessment data. *Clin Chem Lab Med* 2015;53:1725–36.
16. Liu J, Han CH, Loh TP, Badrick T. Detecting long-term drift in reagent lots. *Clin Chem* 2015;61:1292–8.
17. Bais R. Use of capability index to improve laboratory analytical performance. *Clin Biochem Rev* 2008;29:S27–31.
18. De Grande L, Goossens K, Van Uytendaele K, Stöckl D, Thienpont L. The empower project – a new way of assessing and monitoring test comparability and stability. *Clin Chem Lab Med* 2015;53:1197–204.
19. Mackay M, Hegedus G, Badrick T. A simple matrix of analytical performance to identify at assays that risk patients using External Quality Assurance program data. *Clin Biochem* 2016;49:596–600.
20. Tolan NV, Parnas ML, Baudhuin LM, Cervinski MA, Chan AS, Holmes DT, et al. “Big data” in laboratory medicine. *Clin Chem* 2015;61:1433–40.