

The role of EQA in harmonization in laboratory medicine – a global effort

Graham R.D. Jones*^{1,2}

¹Department of Chemical Pathology, SydPath, St Vincent's Hospital, Sydney, NSW Australia

²University of NSW, Sydney, Australia

*Corresponding author: Graham.jones@svha.org.au

Abstract

There are many activities currently being undertaken in the field of laboratory medicine under the broad heading of "harmonization". These include traceability of results to international reference standards, processes to align results from assays where traceability has not been achieved (analytical harmonization) and international or national clinical guidelines based on studies from many parts of the world. Many of these issues are global in nature, with clinical evidence derived from studies performed in all parts of the world and multinational diagnostic companies providing assays worldwide. As with all aspects of medicine, progress can only be assured where there is evidence of effectiveness of the activities. External Quality Assurance (EQA) programs are designed to meet this need. Currently EQA processes have significant limitations in meeting the global needs of the laboratory medicine community. This paper aims to identify the steps that can be taken to allow current and future EQA programs to provide information on global variation in results. It is only by being aware of result differences that steps can be taken to improve performance.

Key words: external quality assurance; metrological comparability; harmonization

Received: April 04, 2016

Accepted: December 27, 2016

Introduction

There are many activities currently being undertaken under the broad heading of "harmonization" in laboratory medicine (1). These can be divided into two main components, the first being activities aimed at improving the metrological comparability of laboratory results and the second being actions based on reducing unnecessary between-laboratory variation in test requesting and reporting. Examples of the latter include the use of common test names, units and reference intervals as well as the development of clinical guidelines to allow application of evidence-based decision making across countries, regions or the world. This paper is focussed particularly on the first component, the comparability of results. It should be noted, however, that while External Quality Assurance (EQA) traditionally addresses analytical quality, the EQA process can equally be applied to other aspects of laboratory activities and can be used to

assess both differences in other factors e.g. units, reference intervals and test names, as well as changes in response to interventions (2-4). The development of clinical guidelines, which include the use of laboratory results, requires an understanding of result variability to ensure recommendations including specific decision points can be validly used in different locations.

Metrological comparability of results

Metrological comparability is generally obtained by traceability to a common reference standard with a valid traceability chain (5). There is however a wide range of terminology in use regarding this notion. Related terminology and concepts include equivalence of measurement, bias, trueness, measurement uncertainty, accuracy and traceability to which must be added the assay properties of

precision and analytical specificity. The term “comparable” (shorthand for “metrologically comparable”) to describe two results, or sets of results, for the same measurand is appropriate as all medical decision making is performed by comparing one or more results from a patient with information derived from other sources (6). The “other sources” may be previous results from the same patient, a population reference interval or a clinical decision point. If the results are not comparable, for example due to a between-method bias, excessive imprecision or different analytical specificity, then the clinical decision may be erroneous. For this paper, I will describe results as comparable if they are suitable for clinical decision-making. This discussion also raises the issue of having a quality standard or analytical performance specification to decide whether results are actually comparable, i.e. fit for purpose for valid clinical decisions.

A key activity aimed at improving the comparability of results is the use of assays traceable to higher order reference materials and methods. While it has been a requirement of the European Union (EU) In-vitro Diagnostics Directive since 1998, this remains an ongoing activity (7). For purchased kit assays, this is a major activity of manufacturers. Other organisations involved in this activity include national measurement institutes, such as the National Institute of Science and Technology (NIST) in the USA and the Institute for Reference Materials and Measurements (IRMM) in Europe and other National Metrology Institutes which produce reference materials, and organisations like the International Federation of Clinical Chemistry and laboratory Medicine (IFCC) which define reference measurement procedures for analytes such as serum enzymes (8). In addition to its general meaning given above, the term “harmonization” is used to describe an alternate process to achieve comparability of results when traceability to a definitive standard is not possible (9). This can be described as analytical harmonization.

It is only the availability of comparable results that allows the development of international clinical guidelines. Such guidelines, e.g. for the KDIGO guidelines for the diagnosis and management of chronic kidney disease and the American Diabetes

Association guidelines for diabetes are based on studies from many parts of the world and the guidelines are only valid if the results from the patient laboratory are comparable to those used in the clinical studies (10,11). Other harmonization activities directly based on comparable results are the development of common reference intervals.

Certainly, no two methods are exactly the same; the issue is whether the difference is important relative to the clinical question. The writers (and readers) of clinical guidelines need to be aware of possible methodological effects. In the absence of results traceable to higher order methods, guideline writers should be aware whether the amount of variation in available methods is likely to be a problem and include a statement relative to this in the guideline. In the absence of traceable results, the applicability of clinical guidelines and development of common reference intervals still depends on the closeness of results from different methods, which must be assessed by appropriately designed EQA.

The concepts listed above have been described as the “six Pillars” of the temple of traceability (12). The first three pillars, certified reference materials, reference methods and reference measurement services, are all well accepted in clinical chemistry and to a lesser extent in other fields of pathology. The Joint Committee for Traceability in Laboratory medicine (JCTLM) provides an on-line database of these higher order references meeting the relevant ISO standards (13). The fourth pillar is reference intervals and decision points, which are traceable to the higher order references. The fifth is “appropriately organised analytical (internal and external) quality control” to ensure assay performance, and the sixth are targets for uncertainty and error of measurement, i.e. whether the closeness of the results meets the clinical need.

As with all aspects of medicine, progress can only be assured where there is evidence of effectiveness of the activities. With regard to result comparability, EQA programs, which are a component of the fifth pillar from Braga above, are designed to meet this need, as is recognised by participation requirements in the clinical laboratory standard

ISO 15189 (14). Currently EQA processes have significant limitations in meeting the global needs of the laboratory medicine community. This paper outlines some of the significant limitations as well as possible future developments in this area to address these limitations.

There are many variables in the design of EQA schemes, which affect the quality of the information that may be derived from them. A framework for some of these properties has been described by Miller (15). Properties identified as varying between programs include the nature of the material, the target assignment procedure, the presence of replicate samples and performance assessment criteria. Under this scheme, the highest ranking EQA program (level 1) is one with verified commutable materials, value assignment by measurement with a reference measurement procedure or comparison with a certified reference material, replicate samples during the program for assessment of within-laboratory precision and method classifications to allow assessment of bias against all participants and a relevant peer group as well as against the reference target. The Miller paper defines six categories with lower rankings assigned to programs with fewer of these desirable features (15).

The global nature of laboratory medicine and the need for international comparability

Laboratory medicine today is a global activity driven by the need for evidence-based medicine and the rise of multi-national in-vitro diagnostics (IVD) manufacturers. Evidence-based medicine implies practice based on research, which demonstrates the value and benefits of the procedure. Research in laboratory medicine is undertaken in all parts of the world. We are unable to apply such research unless the methods in our laboratories produce results that are comparable with those used to produce the results in the research paper. Ideally, EQA can provide confidence that the results from sources anywhere in the world are comparable and thus suitable for such decision-making.

The IVD industry is now dominated by a relatively small number of companies selling equipment and reagents into most countries of the world. These companies may each have a small number of high-volume manufacturing facilities for the making reagents, calibrators and quality control (QC) materials. If there is significant variation in manufacturing quality, this may affect patient care in many different countries. Such variation may be within designed tolerance of the manufacturing process or, occasionally due to failures in such systems. Variation may also be seen in kits from the same company supplied to different locations. While such multinationals are the major players, in the developing world there are many other suppliers of variable size and quality providing reagents and calibrators for routine use.

Global medicine and limitations of current EQA approaches

In contrast to these global aspects of pathology, it is more common for EQA programs to have a majority customer base in their country of origin, although some also have a significant component of international participants. This means that most programs cannot provide direct evidence regarding comparability of results from studies performed in different countries. It also means that problems with a manufacturers' assay identified in one country may not be recognised elsewhere. Additionally the scale of an assay problem may not be able to be estimated, e.g. is it one batch of reagent released in one country, or is affecting all locations where the product is sold.

Geographic limitations for EQA programs are particularly likely when commutable samples are used due to costs and difficulties in obtaining sufficient sample volumes and delays and costs for sample transportation and storage. Programs with a wide international reach more commonly use more highly processed materials, which may affect some aspects of their performance (see below). At this time, there is also no global forum or organisation for discussion or agreement regarding EQA procedures. The European Organisation for Providers of External Quality Assurance in Laboratory

medicine (EQALM) provides this function in a collaborative manner for European countries but no equivalent body operates at a wider level.

Thus at the current time, EQA is not well structured to provide the laboratory community with information about assay performance to ensure global comparability of results. The sections below provide possible approaches to solving this problem.

Global EQA programs

While it is not a commonly performed approach, global EQA programs with commutable material and reference value assignment are possible. The most significant example was the International Measurement Evaluation Program 17 (IMEP 17) project from IRMM, which was reported on in 2003 (16). In this study, two samples with values assigned for 20 analytes were distributed to 1037 laboratories in 35 countries.

Two major international programs, from Randox (RIQAS) and Bio-Rad (Unity), with over 32,000 and 17,000 participants respectively do have a significant global reach. While attention is given to the materials used in these programs they are generally not verified as commutable and do not have reference method value assignment, factors which can limit the utility of these programs to identify some analytical problems. There have been other programs which are widely subscribed, such as the Holt cyclosporine and the CAP commutable sample programs, but these are still most heavily subscribed in the country of origin. Recently there have also been studies covering a number of countries with commutable material and reference method value assignment (17). While the costs of truly international level one EQA programs is significant, IMEP-17 shows that this approach can be adopted.

The importance of commutability

A commutable material is one that demonstrates the same relative response in two or more analytical systems as that shown by native patient samples (18). Commutability is not a general property of a material but rather is an experimentally veri-

fied property for a material based on performance when measured in two or more methods. A material is more likely to be commutable when it is most like a patient sample. With each additional factor that is performed on a sample the chance of non-commutability increases. Such factors include differences in collection devices, delays in handling, different storage tubes and temperatures, additives, spiking, stripping (e.g. with charcoal), lyophilisation and prolonged storage. These, however, are the factors that can often lead to specific benefits such as range of concentrations, extended stability, large volumes to enable large programs and cost control. While EQA samples are rarely validated for commutability for all analytical methods in the presence of all possible interferences, a fresh serum sample with minimal processing has a high likelihood of being commutable.

The primary focus of an EQA program are the participating laboratories, which are assessing the performance of assays in their laboratory. It is relevant to note that these customers pay for the EQA programs and need to see value for their expenditure in this area. For enrolled laboratories, the basic question they are trying to answer might be expressed as "Is my assay functioning the way it is meant to?". Behind this question is the additional question of how is the assay meant to function. For many laboratories, the answer to this question is that it should perform the way a manufacturer intends it to perform. This paradigm is supported in the regulatory environments where any changes to manufacturer's methods may have the effect of changing a kit method into an in-house IVD (19). When comparison data for result interpretation, such as reference intervals or clinical decision points, are derived from studies using the same method this approach has some validity. In this setting, material commutability (and reference method value assignment) are less relevant, and a material which is not fully commutable can still support a laboratory's performance assessment that the method is performing as intended by the manufacturer.

However, material without validated commutability is significantly less useful if the comparator is derived from a different method. Only a commut-

able material can demonstrate between method bias, or lack of such bias, when different methods are in use. Even at the local laboratory level, it can be useful to be aware whether the laboratory down the road, which may also be used by clinicians using your lab, gives results that are comparable to yours.

Reference-method value assignment

If a laboratory wishes to answer the question about accuracy of traceability to an international reference standard, there is no satisfactory alternative to having target values for the EQA material assigned by such methods with commutability being maintained throughout the whole process. In combination with verified commutable samples, or at least with a high likelihood of commutability, it is possible to “close the loop” and directly assess the uncertainties and biases introduced with the calibration hierarchy.

An additional value to the use of higher order reference methods for value assignment is to allow direct comparison of results from different EQA programs in different geographical locations. For example, if a result obtained with a manufacturer’s method is high compared to a reference method result in one country, but not in another, it suggests product variability from the company. If a method is consistently biased in different programs, it suggests a calibration issue.

Performance specifications

If performance of a method in one program is being compared with performance of the same method in a different program, it is necessary to use the same tools for assessment. This depends on assessing the same type of data (e.g. single results or bias assessed from a number samples), at or near the same concentrations using the same performance specifications. Without the same assessment criteria, a method may appear to be biased in one program and not flagged as biased in another program. To this end fixed limits, for example based on biological variation or demonstrated clinical need, may serve better than statis-

tical limits based on spread of the included data (20). The process of developing common performance specifications for EQA has many issues, particularly related to the meaning assigned to the limits (21). In order to achieve commonality in performance specifications there are at three main aspects, firstly the organisational structure within which the work would be done, secondly agreeing on relevant criteria to produce the specifications and thirdly applying those criteria to produce specifications for each measurand.

Method classifications

In order to make valid comparisons using data from different EQA programs it is necessary to have similar method classification systems. There are a number of criteria that can be used to classify methods such as the method principle, the instrument manufacturer, the instrument model or instrument family, the reagent manufacturer, the calibrator, the claimed traceability or any combination of these. If there is a problem with one instrument or one reagent source, this may only be readily apparent in data from different programs if the same classification system is used. Such systems also need to be sufficiently responsive to manage a transition within a manufacturer’s range, e.g. with a method re-standardisation. The most detailed system may be to include lot numbers of reagents and calibrators in the EQA classification scheme to ensure appropriate comparability and to identify possibly faulty products (22). This requirement for a detailed method classification system may lead to difficulties with the complexity of the information required from participants and with displaying on the report. Participants are likely to provide the information if it can provide a benefit to them, and this may be done with the support of local reagent kit suppliers. Reports to participants may be kept at a simple level, however the detailed method data may be displayed in customisable on-line formats such as are available for participants in the Royal College of Pathologists of Australasia Quality Assurance Programs (RCPAQAP) in Australia. The detailed data can be used by EQA providers to “trouble-shoot” aberrant results and in data aggregation amongst different EQA providers.

Combining national and international programs

The cost and difficulty of “level 1” international EQA programs makes these unlikely to be a viable option for routine global use. An alternative may be to enrol a limited number of representative laboratories from multiple countries in a truly global program with commutable materials and reference-method value assignment. These laboratories are then also enrolled in local programs, also using commutable material, along with all relevant laboratories in the country or region. This linked, two-tiered approach could then allow comparison of assay performances in different parts of the world. Such a structured approach could provide a wider reach than programs such as IMEP-17 mentioned above, where there may be a tendency for only better laboratories to be included, giving a falsely reassuring picture of assay performance.

As an example, if one national programme demonstrates a negative drift for a method and another programme a positive drift, there can exist an unacceptable difference between the results although both programmes are within their respective acceptance limits. An overarching international program may identify this issue and lead to a review, for example of product supply in different locations. The key proposal being made is that the data from different EQA programs should be assessed together – an activity that is currently lacking. Expert analysis of the data can decide on the cause and importance of any differences identified.

Oversight/collation of results

As stated above a key limitation in current approaches to EQA is the lack of a truly global approach. In addition to the factors mentioned above, it is also necessary for the information generated to be reviewed at the international level, for example by an appropriate international professional body. For example if raw data from programs with commutable materials and similar method classification could be combined, then a review could assess the overall variation and iden-

tify manufacturers that deviated significantly from other suppliers. Such accumulation of data would be less robust without the use of commutable materials. If there was also reference method value assignment in some of the programs, the deviation from reference targets could be assessed by country as well as manufacturer. As more information is included, e.g. wider range of concentration, recording of lot numbers, then more detailed analysis of the factors affecting global comparability of results can be assessed. Review of this type of data could advise manufacturers and laboratories about current assay performance, identify areas of need for improvement in comparability, and allow advice to researchers, guideline writers and clinicians regarding interpretation of results from different parts of the world.

Conclusions

There is no doubt that laboratory medicine is a global activity. This is true with regard to the generation of interpretive information through clinical trials, the nature of the major diagnostic manufacturing companies and efforts such as the JCTLM to improve result traceability. These actions can all be described as international harmonization, of the results and the information. By contrast, there has been relatively little activity to address EQA issues on a global scale. By way of example there is no global organisation providing leadership and facilitating communication amongst EQA providers.

There are a number of steps that can be taken to improve the situation. A major advance could include further development of global programs. Even within the current practise of many smaller programs, improvements can be made with the use of commutable material, value assignment with higher order references, common data analysis and performance specifications and harmonized method classification. In practice, these will only happen with co-ordinated action amongst EQA programs allowing adoption of common practices and detailed review of the results produced from the many programs currently available. The role of individual laboratories is also important in driving improvements in EQA by select-

ing programs with the characteristics described in this paper. The continued global harmonization of laboratory all activities, specifically including analytical performance, can deliver uniform, evi-

dence-based practice, but this must be verified by EQA which is fit to support this purpose.

Potential conflict of interest

None declared.

References

1. Plebani M. Harmonization in laboratory medicine: the complete picture. *Clin Chem Lab Med* 2013;51:741–51. <https://doi.org/10.1515/cclm-2013-0075>.
2. Jones GRD. Reporting units for therapeutic drug monitoring: a correctable source of potential clinical error. *Med J Aust* 2007;186:420–1.
3. Jones GR, Koetsier SDA. RCPAQAP First Combined Measurement and Reference Interval Survey. *Clin Biochem Rev* 2014;35:243–50.
4. Koetsier S, Jones GRD, Badrick T. Safe Reading of Chemical Pathology Reports. The RCPAQAP Report Assessment Survey. *Pathology* 2016;48:357–62. <https://doi.org/10.1016/j.pathol.2016.02.018>.
5. De Bièvre P, Dybkaer R, Fajgelj A, Hibbert DB. Metrological traceability of measurement results in chemistry: Concepts and implementation (IUPAC Technical Report). *Pure and Applied Chemistry* 2011;83:1873–935. <https://doi.org/10.1351/PAC-REP-07-09-39>.
6. Panteghini M. Traceability, Reference Systems and Result Comparability. *Clin Biochem Rev* 2007;28:97–104.
7. Directive 98/79/EC. In vitro diagnostic medical devices. Official Journal L 331, 07/12/1998 P. 0001–0037. Available at: <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31998L0079:en:HTML>. Accessed 23rd December 2016.
8. Infusino I, Panteghini M. Standardization in Clinical Enzymology. *eJIFCC* 2005;20:3.
9. Miller WG, Myers GL, Gantzer ML, Kahn SE, Schönbrunner ER, Thienpont L, et al. Roadmap for Harmonization of Clinical Laboratory Measurement Procedures. *Clin Chem* 2011;57:1108–17. <https://doi.org/10.1373/clinchem.2011.164012>.
10. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney International Supplements* 2013;3:1–150.
11. American Diabetes Association. Standards of Medical Care in Diabetes. *Diabetes Care* 2016;39:S1–S112.
12. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: Responsibilities and strategies. *Clin Chim Acta* 2014;432:55–61. <https://doi.org/10.1016/j.cca.2013.11.022>.
13. Jones GRD, Jackson CM. The Joint Committee for Traceability in Laboratory Medicine (JCTLM) – its history and operation. *Clin Chim Acta* 2016;453:86–94. <https://doi.org/10.1016/j.cca.2015.11.016>.
14. ISO 15189:2012. Medical laboratories – Requirements for quality and competence. International Standards Organisation, 2012.
15. Miller WG, Jones GRD, Horowitz GL, Weykamp C. Proficiency Testing/External Quality Assessment: Current Challenges and Future Directions. *Clin Chem* 2011;57:1670–80. <https://doi.org/10.1373/clinchem.2011.168641>.
16. Örnemark U, Van Nevel L, Smeyers P, Harper C, Taylor PDP. The International Measurement Evaluation Programme IMEP-17 Trace and Minor Constituents in Human Serum. 2003. European Commission Institute for Reference Materials and Measurements. Available at: https://ec.europa.eu/jrc/sites/default/files/imep17_report_part2.pdf. Accessed 23rd December 2016.
17. Weykamp C, Secchiero S, Plebani M, Thelen M, Cobbaert C, Thomas A, et al. Analytical performance of 17 general chemistry analytes across countries and across manufacturers in the INPUS project of EQA organizers in Italy, the Netherlands, Portugal, United Kingdom and Spain. *Clin Chem Lab Med* 2017;55: 203–11. <https://doi.org/10.1515/cclm-2016-0220>.
18. Miller WG, Myers GL, Rej R. Why Commutability Matters. *Clin Chem* 2006;52:553–4. <https://doi.org/10.1373/clinchem.2005.063511>.
19. National Pathology Accreditation Advisory Council. Requirements for the Development and Use of In-House In Vitro Diagnostic Medical Devices -Third Edition 2014. NPAAC, Australia. Available at: <http://www.health.gov.au/internet/main/publishing.nsf/Content/health-npaac-dhaivd.htm>. Accessed December 23rd 2016.
20. 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. <https://doi.org/10.1515/cclm-2015-0067>.
21. Jones GRD. Analytical performance specifications for EQA schemes – need for harmonization. *Clin Chem Lab Med* 2015;53:919–24.
22. Stavelin A, Riksheim BO, Christensen NG, Sandberg S. The Importance of Reagent Lot Registration in External Quality Assurance/Proficiency Testing Schemes. *Clin Chem* 2016;62:708–15. <https://doi.org/10.1373/clinchem.2015.247585>.