Behind the scenes of EQA - Characteristics, Capabilities, Benefits and Assets of External Quality Assessment (EQA)

Part III - EQA samples

Christoph Buchta[1*], Rachel Marrington[2], Barbara De la Salle[3], Stéphanie Albarède[4], Xavier Albe[5], Tony Badrick[6], Heidi Berghäll[7], David Bullock[8], Christa M. Cobbaert[9], Wim Coucke[10], Vincent Delatour[11], Wolf-Jochen Geilenkeuser[12], Andrea Griesmacher[1,13], Gitte M. Henriksen[14], Jim F. Huggett[15,16], István Juhos[17], Martin Kammel[18,19], Peter B. Luppa[18,20], Piet Meijer[21], Jonna Pelanti[7], Paola Pezzati[22], Sverre Sandberg[23], Michael Spannagl[18], Marc Thelen[24,25], Annette Thomas[26], Heinz Zeichhardt[18,27], Veronica Restelli[28], Lucy A. Perrone[28]

1) Austrian Association for Quality Assurance and Standardization of Medical and Diagnostic Tests (ÖQUASTA), Vienna, Austria;
2) Birmingham Quality (UK NEQAS), part of University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom
3) UK National External Quality Assessment Scheme (UK NEQAS) for Haematology, Watford, United Kingdom
4) Centre Toulousain pour le Contrôle de qualité en Biologie clinique (CTCB), Toulouse, France
5) Co-Chair of the EQALM Virtual Microscopy Working Group; EQALM Virtual Microscopy Working Group Steering Committee Member
6) Royal College of Pathologists of Australasia Quality Assurance Program, St Leonards, Sydney, Australia
7) Labquality, Helsinki, Finland

8) former Director, Birmingham Quality, former President, UK NEQAS

9) Director of the Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Centre, Leiden, The Netherlands; Chair of the IFCC Scientific Division EC and chair of the EFLM Task Force on European Regulatory Affairs

10) Sciensano, Brussels, Belgium

11) Laboratoire National de Métrologie et d'Essais (LNE), Paris, France

12) former Reference Institute for Bioanalytics, Bonn, Germany

13) Central Institute of Clinical and Chemical Laboratory Diagnostics, University Hospital of Innsbruck, Innsbruck, Austria

14) DEKS, Rigshospitalet-Glostrup, Denmark

15) National Measurement Laboratory, LGC, Teddington, United Kingdom

16) School of Biosciences and Medicine, Faculty of Health & Medical Sciences, University of Surrey, Guildford, United Kingdom

17) School of Biosciences and Medicine, Guildford, United Kingdom

18) University of Szeged, Szeged, Hungary

19) INSTAND e.V. Society for Promoting Quality Assurance in Medical Laboratories, Düsseldorf, Germany,

20) GBD Gesellschaft für Biotechnologische Diagnostik mbH, Berlin, Germany

21) Institut für Klinische Chemie und Pathobiochemie, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

22) ECAT Foundation, Voorschoten, The Netherlands

23) SOD Sicurezza e Qualità AOU Careggi Firenze, Italy

24) The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Bergen, Norway

25) Foundation for quality assessment in laboratory medicine (SKML), Nijmegen, the Netherlands
26) Department of Laboratory Medicine, Radboud Medical Center, Nijmegen, the Netherlands
27) Weqas, Cardiff, United Kingdom
28) IQVD GmbH, Institut für Qualitätssicherung in der Virusdiagnostik, Berlin, Germany
29) Canadian Microbiology Proficiency Testing Program (CMPT), Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada

*Corresponding author: Christoph Buchta, Austrian Association for Quality Assurance and Standardization of Medical and Diagnostic Tests (ÖQUASTA), Hörlgasse 18/5, 1090 Vienna, Austria; christoph.buchta@oequasta.at

Running head: Behind the scenes of EQA Part II

Word count: 5704

Tables: none

Figures: 2

Key words: EQA; external quality assessment; PT; proficiency testing; interlaboratory comparison; stakeholders

Statement

Points of view in this document are those of the authors and do not necessarily represent the official position of their affiliated organisations. Where certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible, such identification does not imply recommendation or endorsement by the authors, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Abstract

Providers of external quality assessment (EQA) programs evaluate data or information obtained and reported by participant laboratories using their routine procedures to examine properties or measurands in samples provided for this purpose. EQA samples must offer participants an equal chance to obtain accurate results, while being designed to provide results in clinically relevant ranges. It is the responsibility of the EQA provider to meet the necessary requirements for homogeneity, stability and some other properties of the EQA items in order to offer participants a fair, reliable and technically interesting EQA experience. Thus, the samples are at the heart and in the centre of EQA and its success depends on their quality. This manuscript describes the requirements for EQA samples and the activities of EQA providers to achieve them.

Introduction

This is Part III of a five-part series of articles describing principles, practices and benefits of External Quality Assessment (EQA) of the clinical laboratory. Part I describes historical, legal and ethical backgrounds and properties of individual EQA programs [ref]. Part II deals with key properties of EQA cycles [ref]. Part III is focused on the characteristics of EQA samples. Part IV summarises the benefits for participant laboratories [ref], and Part V addresses the broad benefits of EQA for stakeholders other than participants [ref].

All laboratories enrolled in a particular EQA program or cycle receive material distributed within a similar time period and thus have comparable initial conditions for analysis. To ensure this, the materials used as EQA samples must meet certain properties that, on the one hand, allow their use in EQA and, on the other hand, still simulate patient samples as closely as possible as they are routinely analysed in the laboratory. The first ideally requires homogeneity of samples and stability of the material during storage and transport, and the second ideally requires commutability, a property of sample materials that ensures the ability to evaluate comparability of examination results reported from different measurement procedures (MP) over the intended dynamic range to be measured. However, the MP of different in-vitro diagnostic device (IVD) manufacturers might differ so much that even measuring patient samples on different systems can cause different results. In order to improve this situation, various groups are working on harmonising the laboratory results. For now, it remains the task of the EQA provider to select suitable sample material.

EQA sample materials, types, sources and characteristics

EQA samples can take many forms but ideally should look and behave like actual patient samples so they can be manipulated in the same way as an actual patient sample through the total testing process (TTP). The EQA provider is responsible for the selection of appropriate EQA samples which can be produced by either the EQA provider themselves or by a subcontractor.

Samples can be made from real biological/patient specimens (e.g. serum, plasma, whole blood, urine, sputum, fluid, semen, faeces, saliva, sweat, hair, nasal/nasopharyngeal/cervical swab). These can remain either in their native form or stabilised. Samples can also be formulated using an artificial matrix containing known compounds of interest (e.g. a simulated stool specimen in an artificial matrix spiked with *Salmonella spp*). Samples can also be formulated from a commercially purchased quality control sample, known standards, or other reference or calibration materials. In addition to fluid-based substrates, EQA challenges can also take the form of prepared histological slides, a digital image, a data set for analysis, a paper-based questionnaire/knowledge assessment, or a clinical case scenario. The use of paper and video challenges is a helpful tool for EQA providers in addition to providing 'wet' samples because it allows the provider to assess the pre- and post-analytical phases of the TTP better than a wet sample alone. Paper and video challenges can assess participants' knowledge of other quality management principles and practices such as those regarding reporting, biosafety and document control during the examination process, and pre-analytical knowledge of proper test selection and associated sample preparation with regard to the presented clinical case.

The concentration range of measurands in samples used in the challenges provided in the cycles should reflect the intended purpose of the test in the care pathway as well as the clinical decision

limits and/or suitability of the examination procedure. For quantitative programs, samples at clinically relevant concentrations should be selected to provide an assessment of the potential impact of measurement errors on diagnostic sensitivity and specificity close to clinical decision limits. Where the limit of detection (LOD) directly influences the clinical accuracy (e.g. infectious agent detection), concentrations may be selected that are close to this point to more accurately assess this analytical target, a challenge that could reveal surprising deviations from the manufacturer's specifications [1]. EQA materials can be selected or spiked with known interfering substances to assess if and/or to what extent some in-vitro diagnostic medical devices (IVD-MDs) are affected, providing an indication that these may have perfectable selectivity. While the design of programs or cycles with an analytical focus should include samples covering the full analytical range of the MP, a focus on clinical specificity and sensitivity should rather include samples with concentrations in clinically relevant ranges, especially at the clinical decision limits [2]. The use of patient- and volunteer-derived materials is the preferred choice for EQA material; however, materials may not always be available in sufficient quantity or across the full concentration range required. Often, EQA providers may pool similar donations or enhance the concentration of some measurands by adding exogenous compounds to this material. These changes in the native patient samples can alter the materials, so the performance of the pooled or spiked material should be evaluated in comparison to the unadulterated samples. Using materials as close as possible to the clinical specimen (i.e. commutable) minimises any matrix effect that may cause the material to behave differently between examination methods and/or over the dynamic range, allowing for a more accurate assessment of the IVD-MDs when real patient specimens are examined.

Provision of EQA material for some point-of-care testing (POCT) IVD-MDs such as glucometers, International Normalised Ratio (INR), lipid testing devices and blood gas analysers, is associated with the particular challenge that these devices are intended to analyse

whole or capillary blood. In addition to the instability of whole blood samples, the typically large number of participants in such EQA programs may require the preparation of EQA samples using artificial matrices rather than whole blood, although methods have been described that have been used to prepare sample material that has been evaluated as commutable [3,4]. In addition to greater stability, the use of artificial control material also has the advantage that these samples can offer a wider concentration range of the measurands than samples based on native whole blood. To overcome the limitations of artificial matrices, some studies used EQA materials based on fresh whole blood. An example for glucose POCT resulted in observing much better agreement between different types of glucometer devices compared to what was observed using artificial matrices, suggesting that the latter had perfectible commutability [3]. However, an ideal solution for routine use in EQA for POCT has yet to be found. One approach could be a two-compartment EQA sample system containing whole blood in one compartment and a purified compound in another. The participant could then merge and mix the contents of both chambers for a defined short period of time before the analysis, so that stability issues of the material would be minimised. This approach could be a step towards improving the currently unsatisfactory properties and capabilities of EQA sample materials for IVD-MDs for POCT use.

Selection of starting materials depends on the intended purpose of the samples and the goal of the EQA program. In order to compare the results of individual laboratories only with those of their peers, as in case of EQA categories 5 and 6, homogeneity and stability of the measurands must be verified (Table 1). If samples are intended for harmonisation monitoring and comparing the results of different peer groups (EQA categories 1-4), commutability of the sample material must additionally be verified for each test system included in the EQA program (for more information, see Chapter *Commutability*). In the case of EQA category 1 and 2, the assigned value must additionally be determined using a reference measurement procedure

(RMP) to allow accuracy assessment and standardisation evaluation. As required by the harmonised standard under IVDR, ISO 17511:2020, IVD-manufacturers should adopt and implement the concept of metrological traceability for measurands for which reference methods and reference materials are available, and EQA providers should develop EQA tools that allow them to verify trueness [5]. As the verification of the proper implementation of metrological traceability is highly relevant for disease defining analytes and in particular cases for monitoring treatment, the value assignment of commutable EQA-materials with a reference measurement procedure (RMP) meeting APS requirement (i.e. adequate measurement uncertainty (MU)) is necessary in Category 1 and Category 2 EQA programs that strive for trueness verification (see Table 1). Commutability assessment of EQA materials is necessary for Category 1, 2, 3 or 4 EQA programs.

EQA sample materials, matrices and the measurands contained therein, must be characterised by adequate stability, homogeneity and availability at clinically relevant concentrations and at clinical decision limits in order to meet the respective intended purpose of the EQA activity. While these requirements are sufficient for comparison of individual laboratory results with peers, verification of sample commutability and determination of the target value using a reference method are additionally required for comparison of results of different peer groups and for trueness (absolute bias) assessment. In addition to general characteristics expected of sample materials, it may also be necessary for certain measurands to select the persons who donate EQA sample material on the basis of their ethnicity, as Lipoprotein(a) (Lp(a)) plasma levels and size polymorphism depend on it [6]. For more information on challenging EQA samples for the determination of Lp(a), see chapter "*challenging samples*".

Challenging samples

The clinical usefulness of laboratory test results depends on performance evaluation and detection of errors, including those that affect the sample's integrity and the presence of interferences in that sample. Specimens received in the laboratory often do not reflect the crystal-clear samples used in many EQA programs and may not always be good representatives of clinical specimens received in the laboratory. The three major sample interferences found to affect the accuracy and reliability of most laboratory results are haemolysis, icterus and lipaemia. Each of these situations is a potential source of biological and analytical bias, which ultimately compromises the reliability of measurements of different parameters in routine clinical chemistry examination. Most manufacturers have now included hemolytic/icteric/lipemic indices (HIL) in their test repertoire to identify these interferents in samples, and several EQA providers have now established EQA programs to assess the performance of IVD-MDs for EQA materials with abnormal HIL indices.

Some EQA programs also assess the susceptibility of examination procedures to potential interferents as part of their standard programs. Examples include the assessment of the effects of therapeutic drugs and their metabolites on the performance of other measurands, hook effects in immunoassays, steroid cross-reactivity in immunoassays, effects of ascorbic acid in urinalysis, effects of blood in urine on pregnancy testing performance, the assessment of specificity of examination procedures to samples with mixed viruses, effects of endogenous substances and metabolites on measurands, and the distribution of samples near the limit of detection of the assays.

In addition to interferences in samples, physiologically inhomogeneous measurands can also pose challenges for examination and, in the case of incorrect results, for root cause analysis and therefore they are a welcome challenge in EQA: Lp(a) levels may be under- or overestimated and individuals potentially misclassified for their cardiovascular disease risk, as

i) different immunoassay-based tests detect different proportions of the actual mass of the size-polymorphic measurand [7,8],

ii) conversion of the results from mass to molar units [9], especially in combination with

iii) inferior calibration procedures such as serial dilutions of a single calibrator instead of employing multiple independent calibrators [10],

iv) limited measuring ranges that require dilution of the sample, which in turn can result in a mismatch between sample and calibrator [11], and

v) another heterogeneity affecting the immunoassay result is attributed (but not proven) to the indirect measurement of partially glycosylated Lp(a) [12].

As long as no appropriate reference material or method has been recognized [13], but of course also afterwards, EQA providers can contribute to improving the situation by using targeted EQA samples, not least for educational purposes.

As “challenging samples” are often encountered in the laboratory, they must be included in selected cycles as part of the standard EQA and any conclusions should be fed back to the participants with the report. It is also important that the participant acts on the feedback from the EQA provider. However the performance on such samples may be excluded from the aggregated performance assessment of individual laboratories and/or methods.

Homogeneity and stability

Characterisation of homogeneity of EQA samples and stability of materials is important so that their contribution (as MU) to potential between-laboratory differences can be incorporated when achieving a consensus value and to ensure deviations are reflecting laboratory performance and are not due to differences in the samples. It is obvious that homogeneity and stability tests should only be carried out on the finished and prepared samples after they have

been filled into final containers and, if applicable, lyophilised or subjected to a freeze-thaw cycle, in order to reduce the probability that influences during or after sampling for these tests may alter the characteristics of the samples but remain undetected. The EQA provider is responsible for defining homogeneity and stability of the materials that they distribute and ensuring they are within acceptable parameters. They may do this at two different steps, i) prior to distribution by analysis of a subset of samples (though for some measurands/samples homogeneity and stability do not need to be done for each production if the EQA providers can document that samples are produced in the same way as samples where homogeneity and stability has been proven) and/or ii) post distribution by comparison of imprecision data of current samples against that at a similar concentration from previous cycles, or even better from previous circulations of the same samples (under a different name or sample code to mask such repetition). The advantage of including all the participants' results (usually many times the number of analyses in formal studies) is that it will include any impact of sample transportation.

Procedures applied in preparation of EQA samples may follow the requirements of ISO 33405:2024 on Reference materials - Approaches for characterization and assessment of homogeneity and stability [14]. Ideally, EQA samples should be characterised to the degree of inhomogeneity for each characteristic (measurand) of interest; in practice, the analysis may be limited to the determination of selected characteristics, provided their established chemical or physical relationships to measurands for which they serve as surrogate markers of homogeneity and stability. ISO 17043:2023 refers EQA organisers to ISO 13528:2022 for the statistical analysis of homogeneity and stability [15].

With regard to the stability of materials, a basic distinction must be made between long-term (shelf life), short-term (under "transport conditions") and in-use stability (e.g. stability after reconstitution of lyophilised samples) stability. While "shelf-life" provides information about the period of time during which the properties of EQA materials remain within the tolerable

limit under ideal storage conditions, extreme transport conditions such as delayed delivery and extreme weather, may need to be simulated in stability studies. Post-distribution assessment of between-laboratory agreement includes variations due to transport effects on stability.

Commutability

"Commutability is a property of a reference (or EQA) *material that means results for a reference material [...] and for clinical samples have the same numeric relationship, within specified limits, across the measurement procedures (MPs) for which the reference material is intended to be used. Consequently, a commutable reference material produces a measurement result that is equivalent to the measurement result that would be obtained for a clinical sample with the same concentration of the measurand"* [16]. Commutability has emerged as a key property of i) Certified Reference Materials (CRMs) used in the calibration hierarchy of an end-user examination procedure and ii) EQA materials used as trueness verifiers [17].

All modifications of the matrix of an EQA sample material during its preparation can affect its commutability, like lyophilization, freezing, addition of preservatives, pooling, or spiking with exogenous substances [18]. This can lead to the inability of the EQA material to mimic the behaviour of clinical specimens and to conduct a proper between MPs evaluation. Behaviour in this context could reflect inter methodological biases that occur when comparing clinical samples not being mimicked by the EQA sample or the dynamic range differing between real samples and those used for EQA. At present, commutability in EQA is just starting to emerge and mostly applies for measurands in clinical chemistry. Work is being undertaken on how EQA providers can implement this routinely into their processes [16].

As some examination procedures are more affected than others by matrix effects, the commutability of an EQA material should be evaluated between each pair of examination

procedures, which means that an EQA material may be commutable between some examination procedures but not between others [9,19]. As a result, trueness assessment will be biased in a way that differs across IVD-MDs, leading to an inability to properly estimate both harmonisation and accuracy of the different assays [20]. It has been reported for some measurands that commutability of EQA sample materials proved to be highly heterogeneous, which means that commutability cannot be predicted [21].

The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Working Group on commutability has developed a series of recommendations for assessing commutability [22] according to the difference in bias approach [23] and the calibration effectiveness approach [24] and for correcting the bias caused by non-commutability [25]. In a recent IFCC guideline for commutability assessment in EQA, a practical online tool is presented to the commutability of EQA samples: The criterion for assessing commutability of an EQA sample material between two IVD-MDs is that its result should be within the prediction interval limits based on the statistical distribution of the clinical sample results from the two IVD-MDs being compared. A presupposition for this is that the differences in non-selectivity between the two IVD-MDs being compared are acceptable, the heterogeneity of the measurand is tolerable, and the quantity of the measurand is expressed in molar units [16].

Commutability of EQA materials is necessary in the context of EQA data aggregation. Combining results from various EQA providers may provide a powerful tool to monitor harmonisation of examination procedures in the medical laboratory. Therefore, the International Consortium for Harmonization of Clinical Laboratory Results (ICHCLR) [26] and the European Organization of External Quality Assessment Providers in Laboratory Medicine (EQALM) [27], have joined forces for an initiative called (HALMA) [28]. The acronym stands for HArmonization of measurands in Laboratory Medicine through data Aggregation and aims to collect and aggregate results from different EQA providers that use commutable samples.

The purpose is to evaluate and assess the harmonisation of measurands through aggregated EQA data on an international basis. A working group on commutability works in parallel with the IFCC working group on commutability in metrological traceability on a definition of minimum criteria and evidence to accept that samples used in an EQA Program are, with a high probability, sufficiently commutable to represent examination procedure performance for authentic patient specimens.

Commutability of materials is not binary (i.e. commutable or not) but can range, with some materials being more commutable than others. Although a commutability assessment can have three types of outcomes (commutable, non-commutable and inconclusive), such an assessment provides a quantitative assessment of the non-commutability bias. This makes it possible for the EQA providers to evaluate the degree of non-commutability of a material and decide whether it is suitable or not for the intended purpose. Reaching the conclusion that a given material is commutable means that the non-commutability bias is sufficiently small compared to the analytical performance specification (APS) and/or the clinical application of the MP so that the suitability of the material is not compromised [9]. As EQA materials may be used for different purposes, the level of commutability is not the same for all EQA programs. However, non-commutability between reagent lots may hamper this evaluation and reagent lots should therefore be registered in EQA [29].

Organization of a commutability study for EQA materials includes a number of practical challenges Details on the organisation of a commutability study have been published [16], In brief: At least 30 clinical specimens are obtained, which should be as close as possible to unadulterated clinical specimens, ideally, fresh individual donations. In some cases, this requires specimens from sick donors (usually left over samples). Sourcing fresh single donations is preferable when logistically manageable. An alternative is to use frozen single donations that require demonstrating absence of pooling or freeze/thaw effects on the

concentration of the measurand. The specimens as well as the EQA material should be analysed in multiple replicates within a short time interval and at the same time on all MP and together with the EQA sample(s) whose commutability is to be assessed. Details on how to assess the commutability is given in [16] and a user-friendly free online application on how to do this is available [30]. Be aware that since EQA materials are used for a different purpose than CRMs (Table 1 in [16]), the procedures for assessing commutability of CRMs and EQA sample materials are different [16,31].

While challenging, commutability assessment of EQA materials is critical to appreciate how data analysis can be conducted. When commutability of EQA materials is unknown, results from different peer groups should only be compared with caution. Also, assigning RMP target values to materials of questionable commutability is not desirable as accuracy assessment could be wrong.

The challenge of preparing EQA materials

EQA samples should be such that conclusions can be drawn from the deviations of the measurement results in EQA about the consistency of the results obtained in routine diagnostics with the analytical method to be assessed. To judge the suitability of an EQA sample panel to assess the performance of its participants, the categorisation of Miller et al. is helpful and shown in Table 1 [32]. It should be noted that this categorization does not represent a hierarchical order of superior and inferior materials, but merely a classification according to their possible use.

An essential and defining characteristic of biology is the transience, the temporal change of biological systems and their parts, including humans and the specimens obtained from them. While the limited stability of some measurands in clinical specimens is known and considered

in laboratory diagnostics, EQA samples must be designed to withstand even more challenging storage and transport times and conditions between production and analysis. Therefore, various measures are taken to give EQA samples appropriate properties. The first is cooling or freezing, which is also common for human specimens, but not all EQA samples can be cooled or frozen without destroying or at least impairing the measurands and/or the matrix. If cooling or freezing of a sample material is possible in principle, it must be decided whether to accept the disadvantage of an expensive cooling or freezing transport to the participant, or a possibly restrictive change in the samples during transport. These procedures are also used for patient specimens, but other methods for stabilisation are used for EQA samples. In the procedure of freeze-drying (lyophilization) to extend shelf life or make the material more convenient for transport, the water is removed from the samples, which stops biological decomposition processes. However, only materials that remain stable when freezing can be stabilised by lyophilization, so the advantages of this procedure are limited to those that could also be analysed in samples that are stored and shipped frozen. Stabilisers such as protease inhibitors like sodium dodecyl sulphate, and preservatives (i.e. antimicrobial biocides that kill or inhibit the growth of microorganisms) like sodium azide and guanidine isothiocyanate are also used to stabilise both measurands and matrices in samples.

Not all materials are available in the required concentrations in specimens of healthy donors. To produce materials with clinically relevant concentrations of certain measurands in sufficient quantities for EQA purposes, a matrix from a human donor plasma or serum can be spiked with purified substances. Depending on the substance, production of purified compounds in their pure form is often carried out by e.g., chemical-pharmaceutical processes, by animals, by bacteria in bioreactors, or virus culture methods. Synthetic substances often do not fully correspond to the endogenous compound of human origin. Depending on their specificity, different test systems already detect measurands of human origin to varying degrees. These

differences can be even greater when samples are spiked with exogenous non-human materials. It should be considered that binding of some artificial substances to specific binding proteins in blood may not correspond to authentic substances of human origin. Therefore, spiking EQA materials with exogenous compounds should always be assumed as non-commutability unless proven otherwise and/or a (in)voluntary generation of challenging samples.

Spiking, cooling, freezing, lyophilizing, adding stabilisers or preservatives, but also simply the passage of time - all of these factors can affect commutability of EQA samples [16]. Therefore, the EQA provider must be well familiar with the evaluation of advantages and disadvantages and the selection of the procedures to be used to produce samples.

After all the fundamental obstacles mentioned so far in this chapter have been considered, producing EQA samples begins with selecting the starting material, i.e. the matrix. The starting material can be of human origin (e.g. whole blood, serum, plasma, urine) or produced synthetically. Starting material of human origin can potentially contain pathogens existing in donor blood and thus be infectious. Donor blood or blood components used for EQA purposes may, therefore, be tested like blood for transfusion purposes for the presence of pathogens. Yet laboratories are pointed out to the existing potential infectiousness. Of course, things are different with EQA samples for pathogen detection; In this case, samples with existing or added, but possibly also inactivated, pathogens are produced and labelled as "biohazardous" and sent for examination in EQA cycles. For some EQA samples, substances corresponding to measurands are added to the matrix. This is necessary so that the materials remain unaltered for the time between production, storage and shipping for analysis in the laboratory. The completed EQA sample material is filled in the required volume into suitable containers. These are then sealed or, if necessary, the filled samples are lyophilized before sealing. The stability of formulations has already been determined on their qualification as a sample material; nevertheless, stability tests of individual batches are sometimes performed to demonstrate the

continued suitability of the formulations. For this purpose, relevant measurands are determined in several individual samples and the results are compared with those obtained from samples of the same batch later (e.g. after completing the EQA cycle or at the end of the accepted durability period). There is sufficient stability if the results differ by less than the acceptable maximum. Completed homogeneity tests confirm that it can be excluded with the highest statistical probability that divergent results obtained from different samples of the same batch are attributed to their different content (differences in concentration or absolute content). ISO 33405:2024 describes processes for assessing the homogeneity and stability of reference materials [14]. During and after conducting homogeneity and stability tests, the sample materials are stored under appropriate conditions, i.e. ultra-deep frozen, frozen, cooled or stored under environmental conditions. The samples are stored temporarily at least until the required proof of their homogeneity and, if necessary, their stability has been provided. Many EQA providers also examine the commutability of EQA samples at the time of the batch release. (Figure 1)

The production of EQA materials for several cycles, as well as the joint procurement of materials by several EQA providers, have economic advantages since the characterization of the material is associated with high costs, especially if reference methods are used. Production in bulk works well for highly stable substances such as Cortisol, Sodium, TSH, DNA derivatives of well-established cell lines, etc. Still it would not be possible for whole blood material without the addition of preservatives/stabilising agents. The advantage of having a repeat distribution of the same material is that it is possible to review performance over time, not only at the snapshot of a single EQA cycle. In addition to the economic benefits, another advantage of multiple EQA providers using the same material in their cycles is that it significantly increases the number of laboratories and test systems analysing the same samples at approximately the same time, allowing a more comprehensive comparison of laboratory and

test method performance, even in different geographical regions, and thus also in different framework conditions. Such concerted EQA activities, their benefits and examples are presented in section "EQA providers' networks" in Part V (ref).

Digital samples

In conventional EQA, the provider sends physical samples to laboratories for analysis, and the laboratory sends their findings back for evaluation. An alternative delivery of the 'sample' is digital EQA, where digital items are used instead of physical ones. A digital sample can be for example, a digitalised picture of body fluid or tissue on a slide, a digital cardiogram or pulse curve, DNA sequence(s), lung or heart sound recordings, video of moving objects such as spermatozoids in semen, additionally patient data or examination results. The digital sample is then generally analysed as a conventional sample by using visual or audio inspection or image or data interpretation. It can be distributed through the Internet and shown to laboratories in an interpretable format using a specific software e.g., a virtual microscope for digital smears. However, the whole process (e.g., digitalisation of a physical smear, storing, sharing, and displaying the digital image) requires a sophisticated technical background and infrastructure, but a digital sample has many advantages as compared to a physical one. The same digital sample is shared with all participants, thus guaranteeing a fair and reliable EQA. Digital samples eliminate many problems related to physical samples such as stability, homogeneity and commutability. Finally, the availability of physical samples might be a problem especially in the case of human specimens, because a considerable number of specimens could be necessary for numerous EQA participants, or in specific situations, only a small amount of material can be collected (for example paediatrics specimens). Even more, for "pathological specimens", the only possible way to share them with the EQA participants may be the digital

one. Moreover, digital samples can be used in different surveys even by several EQA providers, enabling and fostering cooperation, standardisation and exchange of expertise. Digital technology offers new tools for laboratories. As an example, in addition to the visual analysis of a digital image, markers and annotations can be placed to identify objects, and quantitative measurements, such as object size or distribution, can be performed. The analytical laboratory analysis workflow can be traced, and in some parts, the pre-analytical flow can be traced; thus, EQA providers can gain insight into the participants' work, getting extra information about the conventional approach. As an illustration, a provider can check which zones of a digitised smear have been searched by a laboratory, but also which zones have not been viewed. Technology allows providers to give fast and precise feedback to laboratories and to offer them advanced and efficient educational support. In addition, the available data collected from the digital analysis of laboratories can be used for data mining and machine learning to improve EQA evaluation or to create intelligent analysis tools for educational or clinical purposes [33].

Digital EQA, using appropriate digital tools, supports collaboration and exchange of information not only for EQA providers and laboratories but also for other parties such as manufacturers and educational organisations. These kinds of interactions are demanded in a globalising (EQA) world. International scientific organisations like EQALM can improve the efficiency of the joint work by identifying differences in the practices of collaborating parties and trying to reduce them through a standardisation process and by making recommendations [34]. To be useful, these recommendations should summarise the knowledge and expertise of different EQA providers and be as consensual as possible. Utilising these recommendations within digital EQA tools will transform them into "living" standards, ensuring they are always up-to-date and scalable, making them easy to adapt and distribute.

Digital imaging may not be a method that laboratories use daily due to implementation challenges such as higher costs and the need to adapt laboratory workflows. In addition the state

of the art of scanning technology may not be sufficient for certain samples requiring the observation of very fine details or moving objects in liquids. However, digital imaging is a constantly evolving technology which will probably solve these drawbacks. At last, an extensive training of laboratory scientists is required to ensure they become proficient, effective and comfortable with this new technology, maintaining high quality standards. Despite these challenges, numerous projects have been initiated and realised to support the use of digital samples in EQA (e.g. [27,33,34]) and for education (e.g. (36–43)). These projects have demonstrated the multiple benefits and new capabilities of this technology.

Conclusion

Following EQA programs and cycles described in previous parts of this series, the requirements for EQA samples and challenges in their preparation were discussed here. (Figure 2). As EQA is developing beyond the traditional sample to challenge the analytical phase within a laboratory, it now also covers pre- and post-examination phases and POCT, and physical EQA items in some areas are being replaced by digital media. Appropriate homogeneity and stability of the measurands and, if applicable, their commutability are fundamental requirements for EQA samples to ensure that all participants have the same chance of obtaining correct results and that any deviations from the target value cannot be attributed to differences in the individual samples.

The higher the requirements for EQA sample materials and the greater the effort required for their characterisation, the more technical equipment, analytical methods and qualified personnel are needed, both for development and production. Although a more comprehensive characterization of the properties of EQA sample materials, beyond measurand stability and sample homogeneity, may soon be common practice for clinical chemistry, new concepts and

approaches are still needed to develop equivalent EQA sample materials for measurands that are more unstable and/or not represented as mass per volume, such as they exist in haematology, coagulation diagnostics or infection immunology. Ultimately, the aim is to find a balance between the acceptable deviation of EQA sample properties from the native sample material of an individual patient and the acceptable cost of production, which is of course also reflected in the sample costs.

Author contributions:

Christoph Buchta: Conceptualized this review, wrote the manuscript draft, edited and critically reviewed the manuscript; Rachel Marrington: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Barbara De la Salle: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Stéphanie Albarède: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Xavier Albe: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Tony Badrick: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Heidi Berghäll: Provided scientific advice, wrote, edited and critically reviewed the manuscript; David Bullock: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Christa M. Cobbaert: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Wim Coucke: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Vincent Delatour: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Wolf-Jochen Geilenkeuser: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Andrea Griesmacher: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Gitte M. Henriksen: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Jim F. Huggett: Provided scientific advice, wrote, edited and critically reviewed the manuscript; István Juhos: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Martin Kammel: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Peter B. Luppa: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Piet Meijer: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Jonna Pelanti: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Paola Pezzati: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Sverre Sandberg: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Michael Spannagl: Provided scientific advice,

wrote, edited and critically reviewed the manuscript; Marc Thelen: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Annette Thomas: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Heinz Zeichhardt: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Veronica Restelli: Provided scientific advice, wrote, edited and critically reviewed the manuscript; Lucy A. Perrone: Provided scientific advice, wrote, edited and critically reviewed the manuscript. The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Funding sources:

None declared.

Acknowledgements

The authors wish to express their gratitude to Anna Malikovskaia for her support with the management of the extensive bibliography of this five part paper series.

References:

1. Bigio J, MacLean ELH, Das R, Sulis G, Kohli M, Berhane S, et al. Accuracy of package inserts of SARS-CoV-2 rapid antigen tests: a secondary analysis of manufacturer versus systematic review data. Lancet Microbe 2023;4:e875–82.

2. James D, Ames D, Lopez B, Still R, Simpson W, Twomey P. External quality assessment: best practice. J Clin Pathol 2014;67:651–5.

3. Bukve T, Sandberg S, Vie WS, Sølvik U, Christensen NG, Stavelin A. Commutability of a Whole-Blood External Quality Assessment Material for Point-of-Care C-Reactive Protein, Glucose, and Hemoglobin Testing. Clin Chem 2019;65:791–7.

4. Stavelin A, Sandberg S. Essential aspects of external quality assurance for point-of-care testing. Biochem Med (Zagreb) 2017;27:81–5.

5. ISO 15189:2022 Medical laboratories – requirements for quality and competence. Geneva, Switzerland: International Organization for Standardization (ISO); 2022

6. Enkhmaa B, Anuurad E, Berglund L. Lipoprotein (a): impact by ethnicity and environmental and medical conditions. J Lipid Res 2016;57:1111-25.

7. Reyes-Soffer G, Ginsberg HN, Berglund L, Duell PB, Heffron SP, Kamstrup PR, et al. Lipoprotein(a): A Genetically Determined, Causal, and Prevalent Risk Factor for Atherosclerotic Cardiovascular Disease: A Scientific Statement From the American Heart Association. Arterioscler Thromb Vasc Biol. 2022;42:e48-60.

8. Ruhaak LR, Cobbaert CM. Quantifying apolipoprotein(a) in the era of proteoforms and precision medicine. Clin Chim Acta 2020;511:260–8.

9. Dikaios I, Althaus H, Angles-Cano E, Ceglarek U, Coassin S, Cobbaert CM, et al. Commutability Assessment of Candidate Reference Materials for Lipoprotein(a) by Comparison of a MS-based Candidate Reference Measurement Procedure with Immunoassays. Clin Chem 2023;69:262–72.

10. Dati F, Tate JR, Marcovina SM, Steinmetz A, International Federation of Clinical Chemistry and Laboratory Medicine, IFCC Working Group for Lipoprotein(a) Assay Standardization. First WHO/IFCC International Reference Reagent for Lipoprotein(a) for Immunoassay--Lp(a) SRM 2B. Clin Chem Lab Med 2004;42:670–6.

11. Kronenberg F. Lipoprotein(a) measurement issues: Are we making a mountain out of a molehill? Atherosclerosis 2022;349:123–35.

12. Szarek M, Reijnders E, Jukema JW, Bhatt DL, Bittner VA, Diaz R, et al. Relating Lipoprotein(a) Concentrations to Cardiovascular Event Risk After Acute Coronary Syndrome: A Comparison of 3 Tests. Circulation 2024;149:192–203.

13. Ruhaak LR, Romijn FPHTM, Begcevic Brkovic I, Kuklenyik Z, Dittrich J, Ceglarek U, et al. Development of an LC-MRM-MS-Based Candidate Reference Measurement Procedure for Standardization of Serum Apolipoprotein (a) Tests. Clin Chem 2023;69:251–61.

14. ISO 33405:2024 Reference materials — Approaches for characterization and assessment of homogeneity and stability. Geneva, Switzerland: International Organization for Standardization (ISO); 2024

15. ISO 13528:2022 Statistical methods for use in proficiency testing by interlaboratory comparison. Geneva, Switzerland: International Organization for Standardization (ISO); 2022

16. Sandberg S, Fauskanger P, Johansen JV, Keller T, Budd J, Greenberg N, et al. Recommendations for Setting a Criterion and Assessing Commutability of Sample Materials Used in External Quality Assessment/Proficiency Testing Schemes. Clin Chem 2023;69:1227–37.

17. Miller GW, Greenberg N, Budd J, Delatour V. The evolving role of commutability in metrological traceability. Clin Chim Acta 2021;514:84–9.

18. Badrick T, Miller WG, Panteghini M, Delatour V, Berghall H, MacKenzie F, et al. Interpreting EQA - Understanding Why Commutability of Materials Matters. Clin Chem 2022;68:494–500.

19. Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, Jansen R. Commutability assessment of potential reference materials using a multicenter split-patient-sample between-field-methods (twin-study) design: study within the framework of the Dutch project “Calibration 2000.” Clin Chem 2002;48:1520–5.

20. Delatour V, Clouet-Foraison N, Jaisson S, Kaiser P, Gillery P. Beware of Noncommutability of External Quality Assessment Materials for Hemoglobin A1c. Clin Chem 2020;66:390–1.

21. Delatour V, Clouet-Foraison N, Jaisson S, Kaiser P, Gillery P. Trueness assessment of HbA 1c routine assays: are processed EQA materials up to the job? Clin Chem Lab Med 2019;57:1623–31.

22. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design. Clin Chem 2018;64:447–54.

23. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group Recommendations for Assessing Commutability Part 2: Using the Difference in Bias between a Reference Material and Clinical Samples. Clin Chem 2018;64:455–64.

24. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC Working Group Recommendations for Assessing Commutability Part 3: Using the Calibration Effectiveness of a Reference Material. Clin Chem 2018;64:465–74.

25. Miller WG, Budd J, Greenberg N, Weykamp C, Althaus H, Schimmel H, et al. IFCC Working Group Recommendations for Correction of Bias Caused by Noncommutability of a Certified Reference Material Used in the Calibration Hierarchy of an End-User Measurement Procedure. Clin Chem 2020;66:769–78.

26. International Consortium for Harmonization of Clinical Laboratory Results (ICHCLR). Available from: https://www.harmonization.net/

27. European Organization of External Quality Assessment Providers in Laboratory Medicine (EQALM). Available from: https://eqalm.org/working-groups/virtual-microscopy

28. EQALM [Internet]. HALMA - EQALM. Available from: https://eqalm.org/cooperations/halma

29. 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.

30. Sandberg S, Fauskanger P, Johansen J, Keller T, Budd J, Greenberg N. Commutability assessment of external quality assessment material. Available from: https://www.noklus.no/en/a-practical-tool-for-commutability-evaluation-of-external-quality-assessment-material/

31. Miller WG, Keller T, Budd J, Johansen JV, Panteghini M, Greenberg N, et al. Recommendations for Setting a Criterion for Assessing Commutability of Secondary Calibrator Certified Reference Materials. Clin Chem 2023;69:966–75.

32. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency Testing/External Quality Assessment: Current Challenges and Future Directions. Clin Chem 2011;57:1670–80.

33. Juhos I. Present and future possibilities of virtual EQA. Clin Chim Acta 2019;493:S749.

34. Albe X, Juhos I, Albarede S, De La Salle B, Geilenkeuser WJ, Maria E, et al. Digital Blood Smear Analysis EQA Scheme Guidelines. In. Available from: https://eqalm.org/component/rsfiles/download-file/files?path=working-groups%252Fvirtual_microscopy%252FDBSAGuidelines2021.pdf&Itemid=101

35. Burthem J, Brereton M, Ardern J, Hickman L, Seal L, Serrant A, et al. The use of digital "virtual slides" in the quality assessment of haematological morphology: results of a pilot exercise involving UK NEQAS(H) participants. Br J Haematol. 2005 Jul;130(2):293–6.

36. Fraunhofer IIS [Internet]. [cited 2023 Jun 30]. Available from: https://www.iis.fraunhofer.de/en/ff/sse/health/medical-image-analysis/mikroskopie/parasiteweb.html

37. Hutchinson C, Brereton M, Adams J, De La Salle B, Sims J, Hyde K, et al. The Use and Effectiveness of an Online Diagnostic Support System for Blood Film Interpretation: Comparative Observational Study. J Med Internet Res. 2021 Aug 9;23(8):e20815.

36. University of Leeds [Internet]. [cited 2023 Jun 30]. Virtual Pathology at the University of Leeds. Available from: https://www.virtualpathology.leeds.ac.uk/

38. American Association for Anatomy [Internet]. [cited 2023 Jun 30]. Available from: https://www.virtualmicroscopydatabase.org/

39. University of Tasmania [Internet]. 2023. Available from: https://vmc.utas.edu.au/

40. Ahmed L, Seal LH, Ainley C, De la Salle B, Brereton M, Hyde K, et al. Web-Based Virtual Microscopy of Digitized Blood Slides for Malaria Diagnosis: An Effective Tool for Skills Assessment in Different Countries and Environments. J Med Internet Res. 2016 Aug 11;18(8):e213.

41. Lakhtakia R. Virtual Microscopy in Undergraduate Pathology Education: An early transformative experience in clinical reasoning. Sultan Qaboos Univ Med J. 2021 Aug 29;21(3):428–35.

42. Qing J, Cheng G, Ni XQ, Yang Y, Zhang W, Li Z. Implementation of an interactive virtual microscope laboratory system in teaching oral histopathology. Sci Rep. 2022 Mar 31;12(1):5492.

43. Somera Dos Santos F, Osako MK, Perdoná G da SC, Alves MG, Sales KU. Virtual Microscopy as a Learning Tool in Brazilian Medical Education. Anat Sci Educ. 2021 Jul;14(4):408–16.

Abbreviations

APS	Analytical performance specification
CRM	Certified reference materials
EQA	External Quality Assessment
EQALM	European Organization of External Quality Assessment Providers in Laboratory Medicine
HALMA	HArmonization of measurands in Laboratory Medicine through data Aggregation
HIL	Hemolysis, icterus, lipemia
ICHCLR	International Consortium for Harmonization in Laboratory Medicine
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IVD	In-vitro diagnostic
IVD-MD	In-vitro diagnostic medical device
LOD	Limit of detection
Lp(a)	Lipoprotein (a)
MP	Measurement procedure
POCT	Point-of-care testing
RMP	Reference measurement procedure
TTP	Total testing process

Table1: Different types of EQA programs offer different evaluation capabilities

EQA program/cycle design				Evaluation capability			
EQA Category	Commutable materials	Assigned values	Replicate measurements	Precision	Absolute bias	Compare results from different peer groups	Compare individual laboratory results with peers
1	Yes	RMP	Yes	Yes	Yes	Yes	Yes
2	Yes	RMP	No	No	Yes	Yes	Yes
3	Yes	Consensus	Yes	Yes	No	Yes	Yes
4	Yes	Consensus	No	No	No	Yes	Yes
5	No	Consensus	Yes	Yes	No	No	Yes
6	No	Consensus	No	No	No	No	Yes

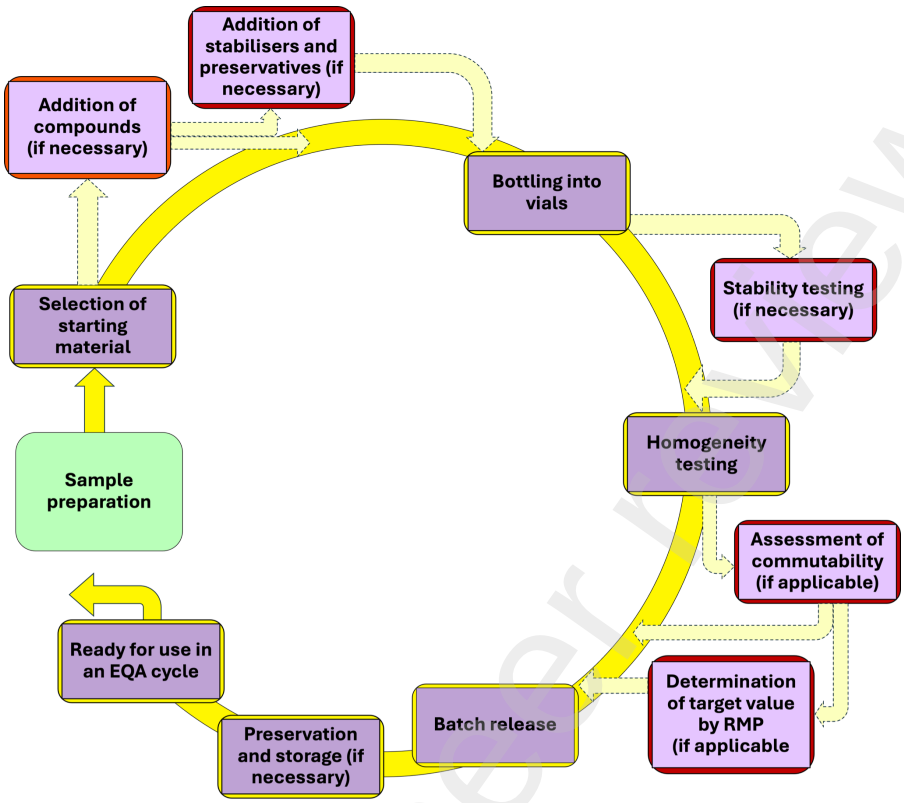
Legend:

(adapted from [32]). According to Miller et al., there are different types of EQA programs:

- Category 1 and 2 EQAs aim at evaluating both standardisation (metrological traceability) and harmonisation of clinical laboratory results (comparability between measurements obtained using assays from different platforms). This requires having commutable materials with assigned values determined using a reference method.
- Category 3 and 4 EQAs aim at evaluating harmonisation of clinical laboratory results. This requires having commutable materials. Assigned values can be determined using consensus means (either all laboratories trimmed mean or all MP consensus mean.
- Category 5 and 6 EQAs aim at verifying the correct implementation of an IVD-MD according to manufacturer's specifications by comparing results obtained in an individual laboratory against those obtained in other laboratories from the same peer group, i.e. using the same method or analytical platform. As materials commutability is unknown, results from different peer groups cannot be compared and results harmonisation cannot be evaluated. As assigned values are determined using consensus means, evaluating results accuracy is not possible.

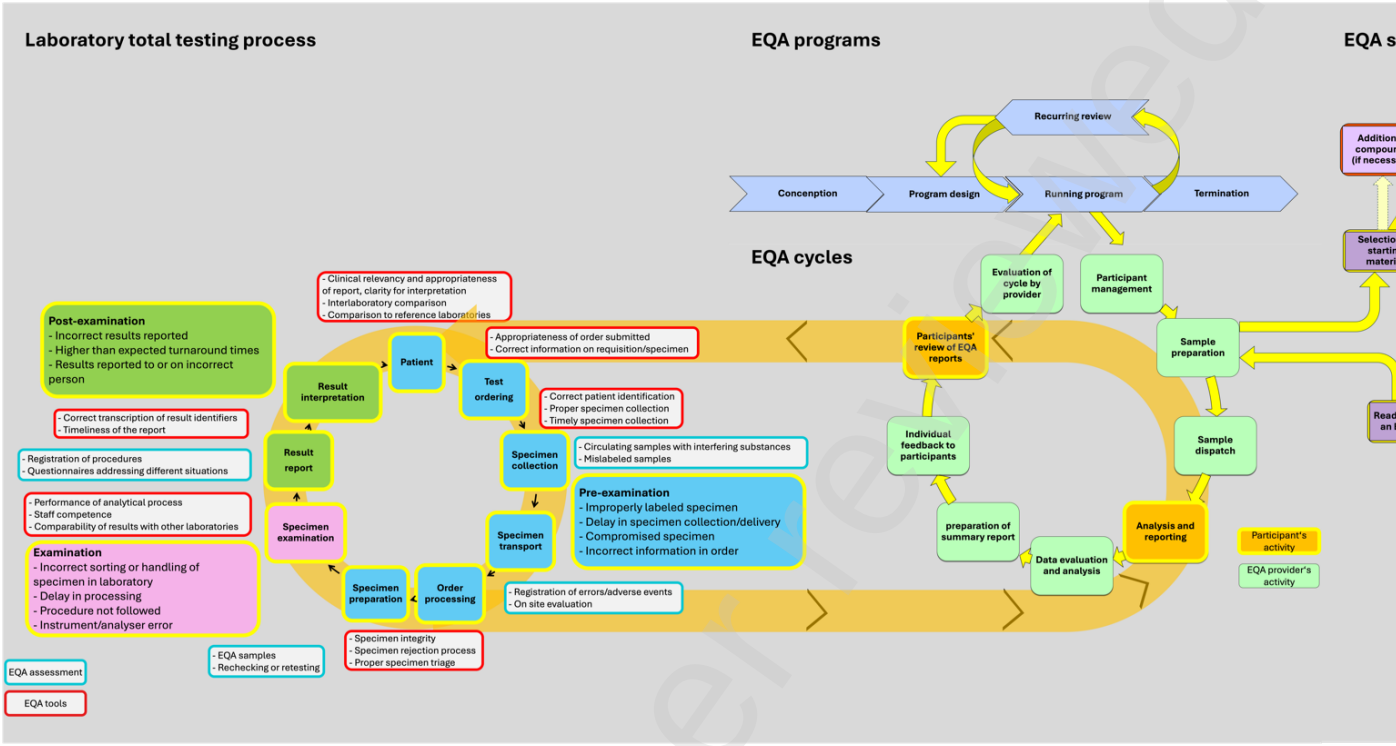
When replicates are performed (as in EQA Categories 1, 3, and 5), precision can be estimated.

Figure 1: EQA sample preparation



Legend: The preparation begins with the determination of the intended use of the individual proficiency test samples, the selection of suitable starting materials and, if necessary, the addition of compounds, stabilisers and preservatives. After portioning and filling into suitable containers, the homogeneity and, if necessary, stability and commutability (EQA Categories 1-4) of the samples are tested; for samples intended to be used in EQA Categories 1 or 2, target values are determined by a RMP. If the results of the quality controls meet the requirements, the batch is released for use in EQA.

Figure 2: Laboratory total testing process, EQA programs, cycles and sample preparation



Legend: Relationship of the laboratory total testing process, EQA cycles, EQA programs, and the preparation of samples used in EQA