Accreditation and Quality Assurance
https://doi.org/10.1007/s00769-024-01582-1

REVIEW

Comparisons of real versus synthetic proficiency testing items

Finlay MacKenzie[1] · Rachel Marrington[1]

Received: 11 January 2024 / Accepted: 26 February 2024
© The Author(s) 2024

Abstract
Both real and synthetic materials are regularly used by providers of external quality assessment and proficiency testing schemes, and there are, in most cases, justifiable reasons for the choice made. This article focusses on the field of clinical biochemistry and discusses the different properties of a material and the benefits and limitations of using real or synthetic material. However, the overarching principles should be able to be applied to all sectors within the laboratory. Whilst genuine/real material would appear to be the preferred matrix, this is not always practicable, and synthetic material may be a suitable alternative. Synthetic material covers a wide range of material, be it that the material is 100% artificial to real material being used as a 'base' item, which is manipulated either by the addition of further real material from a different source or by the addition of exogenous analyte. A number of real-life cases are presented to demonstrate the impact of material matrix, storage conditions, volume, and interferences.

Keywords EQA · PT · Commutability · Homogeneity · Matrix

Introduction

External quality assessment (EQA) or proficiency testing (PT) is at the core of quality assessment, standing alongside internal quality control (IQC) and a robust quality management system [1]. The terms EQA and PT are often used interchangeably, but both share the same principles where sample(s) are distributed to participants. We will use the term EQA/PT here unless there are any specific comments relevant to only EQA or PT. The participant analyses the sample and reports results. Data are then compared to a target/assigned value, and that laboratory's performance is then reported back. The term EQA is mostly used within medical laboratories where it has more of an educational element and may look at all phases of the testing cycle.

Although the broad concepts of EQA/PT are deemed to be well established, there are, despite the guidance of ISO/IEC 17043:2023 [2], no two EQA/PT providers who operate their services identically. Even taking into account the way that different customers/end-users are defined and how they have their services mapped onto different standards, there remains a multitude of approaches used by EQA/PT providers. Though there has been pressure to harmonise, it is unlikely that there will be a one-size-fits-all as different EQA/PT providers have different objectives with the services that they provide. There is a concern that legislation might fuel a race to the bottom and inadvertently remove the nuance and subtilties that distinguish scientifically led comprehensive programmes from the merely just adequate ones.

There are a number of critical aspects that EQA/PT providers need to have available in order to offer a service—a source of material, a means of distributing the material and a method of statistical data analysis of reported results which can then be conveyed back to the participant. According to ISO/IEC 17043:2023, EQA/PT items should 'usually match the type of items or materials encountered in routine laboratory activities' [2]. ISO 15189:2022 states that the EQA programme(s) selected by the laboratory shall, to the extent possible, provide samples that mimic patient samples for clinically relevant challenges [3]. Though this implies 'real' EQA/PT items are preferred, the practicalities, as discussed below, may mean this is impractical, and 'synthetic' EQA/PT items are required. For the purposes of this article, we are using the following conventions:

Real EQA/PT item—Unadulterated material from a single source

✉ Rachel Marrington
rachel.marrington@uhb.nhs.uk

[1] Birmingham Quality (UK NEQAS), University Hospitals Birmingham NHS Foundation Trust, PO Box 3909, Birmingham B15 2UE, UK

Published online: 22 March 2024

Synthetic EQA/PT item—Manipulated material

Synthetic covers a spectrum of items, ranging from the manipulation of genuine material (*e.g.* pooling different sources, addition of exogenous analyte) to the preparation of a matrix that is completely unrelated to the item that is measured in routine laboratories. One also needs to take into account whether the EQA/PT provider is preparing the specimens themselves or subcontracting production to a third party (*e.g.* purchasing control material from commercial companies). The latter may have limitations in that the EQA/PT provider may not be fully aware of the processing that the material they are using has undergone.

The test that most people associate with health care is the so-called 'blood test' which is often interpreted as a single test that does everything, but is in fact thousands of different measurands. Even the term 'blood test' does not fully describe the process, which can be broken down into (1) pre-examination, (2) examination, and (3) post-examination. Although laboratories have most control of the second stage, they do have responsibility for the other two components. A good EQA programme should include pre- and post-examination elements, and the effectiveness of this must, in some way, relate to the quality of the EQA item in the first place. Laboratories accredited to ISO 15189:2022 are required to consider pre- and post-examination components in their quality assessment [3, 4].

This article is drawn from the in-person lecture delivered at Eurachem's 10th Workshop on Proficiency Testing in Analytical Chemistry, Microbiology, and Laboratory Medicine, Windsor, 2023. It is largely clinical chemistry based, but many of the concepts can be applied to the wider EQA/PT arena. Science, whether it be laboratory medicine based or that used in construction materials, is continuously evolving, and change and innovative solutions have to be accommodated in the wider framework.

Table 1 shows the key properties of EQA/PT items and how they apply, at a high level, to real and synthetic EQA. A number of these are discussed in more detail in the sections below.

Matrix/sample types

Clinical laboratories handle a variety of sample types ranging from serum, plasma, whole blood, urine, faeces, fluid, saliva, sweat, and many more. Though an assay should be designed to be specific for a particular analyte, it may use other components within the sample matrix as part of the assay architecture. An example of which are the enzymes AST and ALT. Pyridoxal-5-phosphate (PLP) is a cofactor for both enzymes. This converts the inactive apoenzyme into the active holoenzyme which is then able to catalyse the transfer of amino groups. PLP is the active form of Vitamin B6. Some patients are more likely to be deficient—patients with chronic liver disease, alcoholism, or those with increasing age. This means that unless an AST/ALT assay is used with added PLP (as recommended by the IFCC), it is likely that the enzyme activity levels will be under-reported [5]. It is important that the enzyme activity should be reflecting liver damage and not just nutritional status. In 2023, only approximately 35% of laboratories are reporting results to the UK NEQAS for Clinical Chemistry programme using an IFCC enzyme method [6]. This is valuable information that is provided to the laboratory about the performance of their method/assay compared with other methods. Well-designed EQA will show this for all measurands. The example shown here is particularly relevant due to the prevalence of Vitamin B6 deficiency in the UK [7], and that both AST and ALT are used in a calculated parameter called Fibrosis-4 index (FIB-4).

Table 1 Table to show an overview of how real and synthetic materials differ for a range of properties of a material distributed through EQA/PT programmes.

Property	Real	Synthetic
Matrix/sample types	Available with limitations	Usually need to start with biological base
Storage	Need to use immediately	May have a better shelf life
Commutability	Implicit commutability	Commutability needs to be proven
Concentration ranges	Limited concentration ranges	Wide and challenging concentrations possible
Volume of sample	Volume constraints	Potentially unlimited volumes
Homogeneity	Needs intervention	Needs intervention
Stability	May be limited	Preservatives can be added
Challenging specimens	Difficult-to-source challenging specimens	Construct, within reason, any specimen required
Interferences	Difficult-to-source specimens containing or with knowledge of interferents	Construct, within reason, any specimen required
Financial considerations	Could be minimal if remnant material is donated, but conversely could be high if purchased	Potentially more predictable costings
Transportation	May be limited by stability	Preservatives can be added to extend transportation ability

Binding proteins in serum/plasma can also influence how much of an analyte is available to be ‘captured’ by an antibody on an assay and so can markedly alter the apparent concentration. An example of which is serum progesterone which is now usually measured using a non-isotopic immunoassay (historically this would have been performed using a radio immunoassay (RIA)). We have shown that you can get different results for Siemens assays compared to other manufacturers immunoassays whether the sample is native endogenous, spiked by exogenous pure hormone, or native endogenous serum from pregnant ladies (Fig. 1) [8]. It is hypothesised that the differences are due to the increase in concentration of cortisol-binding globulin (CBG) during pregnancy. CBG binds progesterone, and assay systems require progesterone to be ‘released’ from the CBG before it can be measured and so be in its ‘free’ state; therefore, potentially not all assay systems are effective at this part of the process.

Sample composition can also be affected by the sex of the donor in terms of concentrations of sex hormones present and other binding proteins, etc. Figure 2 shows the relative method biases for serum samples composed of off-the-clot pooled male or female serum, either at endogenous concentrations of cortisol or where cortisol has been added, for four of the major manufacturers. The target/assigned value to which each manufacturer's method mean is compared to is a candidate for reference method cortisol result. The data are from 2021–2023. Though there are concentration dependent biases, the main point to note is that for two manufacturers, the Siemens Atellica and Beckman DxI, there are clear differences in method bias between male and female specimens/pools.

Urine, pleural fluid, ascitic fluid, saliva, and sweat are all aqueous based and can be thought of as being easier to prepare as synthetic materials [9–11]. However, each has the same issue of being present in the human body at a particular pH with other components present including any relevant metabolites. Different disease states may alter this composition; therefore, if fully synthetic material is used, the manufacturer needs to be aware of the limitations of what they are producing. There may be a place for these materials for educational and research purposes, but if they are being used to make specific decisions about an analytical process, then the limitations of production do need to be taken into consideration.

Analytes found in faecal material are increasingly being used in the clinical arena, and so providing samples to laboratories in this matrix is becoming expected and considered more routine. There are fully synthetic alternatives, usually based on looking for the presence or the absence of a substance that can be added in different concentrations. The synthetic part is mainly only to give the ‘look’ of a faecal sample and is a ‘carrier’ for the analyte of interest. This is not possible for all analytes where the analyte of interest has not been isolated in sufficient concentrations in the correct form to be representative of clinical specimens. For example, calprotectin.

Whole blood specimens are widely used in haematology, but less so within routine biochemistry. However, there is a requirement for analytes such as HbA1c and, more recently, for the provision of EQA/PT material for point-of-care testing (POCT). Whole blood is more difficult to handle. Firstly, the red cells have a limited lifespan before they will lyse (unless stabilisers are added), and, due to blood group incompatibility, it can be more difficult to pool together large volumes without agglutination occurring. Red cells are very fragile; therefore, this sample material has to be handled carefully to avoid cell lysis.

Storage

Freezing material allows greater flexibility for the EQA/PT provider for raw material and any subsequent prepared pool. Freezing often extends the shelf life of materials (an exception would be for whole blood as the red cells lyse on

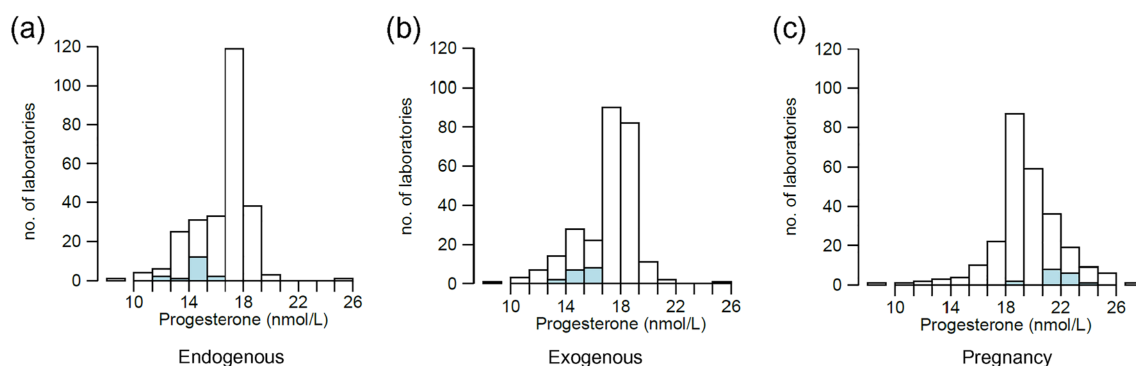

Fig. 1 Histogram of serum progesterone results for three different matrix types, from the same distribution (August 2022). The shaded areas are results from the Siemens Atellica

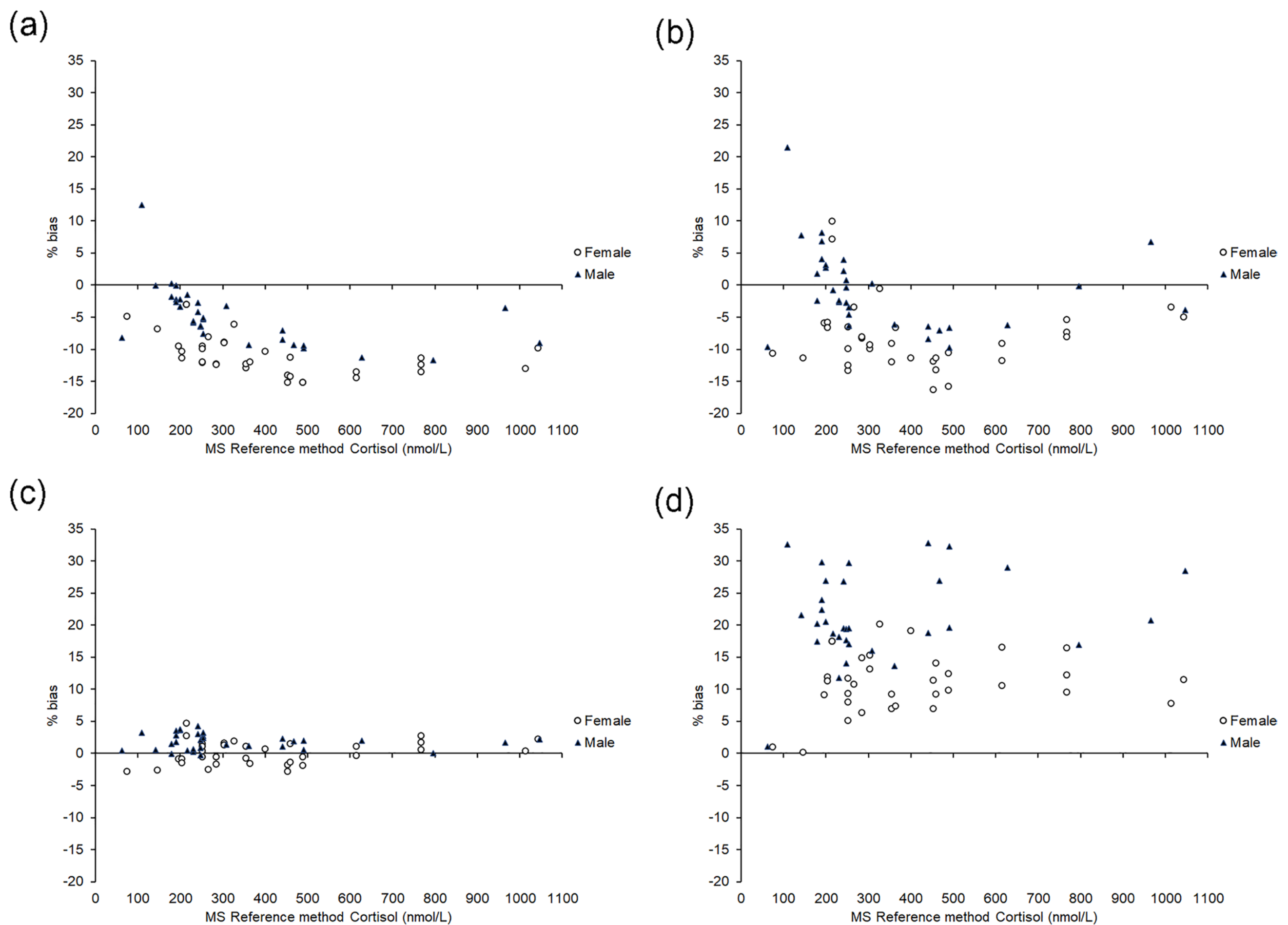

Fig. 2 Plots of serum cortisol specimen % bias for four different methods relative to a candidate reference method for serum cortisol **a** Abbott Alinity, **b** Beckman DxI, **c** Roche Cobas, and **d** Siemens Atellica. The data are from 2021–2023 and are split by patient sex

thawing), but it is important to be aware of the impact of single or repeated freeze/thaw processes. Lipids and lipoproteins are severely impacted by freeze/thaw processes [12]. As lipids and lipoproteins are present in all serum matrices, this needs to be considered, and it is generally advised to minimise the number of freeze–thaw cycles that any material is subjected to. Cramb et al. undertook a comparison between EQA and clinical specimens looking specifically at lipid analysis and concluded that the freeze/thaw process does influence individual manufacturer results (to different degrees) when compared to a reference method [13]. Further work showed improved precision between users when fresh material was distributed, from a single donor, compared to pooled frozen material. Figure 3 shows the imprecision profile over a one-year time period, for HDL-cholesterol for all specimens distributed through the UK NEQAS for Lipid Investigations programme (a) frozen and pooled material and (b) fresh, single donations. The open circles represent the relative coefficient of variation (%CV) between all laboratory methods, whereas the filled circles are specifically for Roche reagents. Though the within-method (imprecision) is consistent and generally less than 5% for the individual method both on frozen, pooled material and fresh, single-donation material, the imprecision between different methods is much wider on frozen, pooled material. Large differences between methods are one reason why a method mean has to be used as a target/assigned value; however, this does mean that the data cannot be used to give any indication of how accurately a method is performing.

There are different protocols available for thawing liquids: keep it at room temperature or warm it in a water bath. These should not be used interchangeably, and material should not be left at room temperature for extended periods of time as it is of biological composition with a defined period of stability. It is often very counter-intuitive that such minor differences in protocols can have an impact on the measurand of interest.

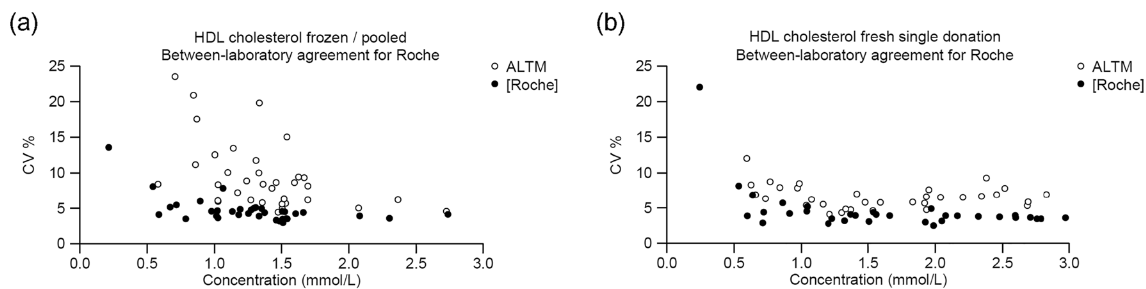

Fig. 3 Between-laboratory agreement plots for two different matrices for HDL-cholesterol. **a** Regular pooled frozen sera and **b** individual fresh single donations. The data are from 2015, and an example is shown here for the imprecision for Roche reagents (filled circles) and the imprecision for the all laboratory trimmed mean [ALTM] (open circles)

Commutability

Commutability is a property of a reference material (RM) that relates to the closeness of agreement between results for that 'material' and results for clinical samples when measured by two measurement procedures [14]. At its most brutally simple, any international standard, Diagnostic Kit Calibrator or EQA material must behave in the same manner as the clinical samples that the procedure is being asked to measure. Superficially, this is a straightforward task, but the practical considerations make this both a logistical and expensive challenge. It is eminently possible, but historically, commutability studies have not been carried out by manufacturers, RM producers, and EQA material producers (whether subcontracted or in-house). There is one major caveat; you need to have assays that are selective for the analyte in question to allow such studies to be undertaken [15]. For example, it is well known for decades that the compensated kinetic Jaffe method is not specific for creatinine and despite standardisation, laboratories still choose to use this inferior assay [16, 17]. Clinical specimens from patients give different results with different methods [18]. No commutability experiment would ever be able to show even the most perfect 'material' was commutable in the case of the compensated kinetic Jaffe assay, as 'selectivity' is a prerequisite for commutability.

Whilst commutability has always been considered as part of the wider 'uncertainty' envelope by EQA/PT providers, it has only relatively recently been placed under such intense scrutiny. The IFCC Commutability in Metrological Traceability Working Group (WG-CMT) has written some general background and guidance on commutability [14, 19, 20] and has followed this up with focussed papers on both EQA materials [22] and other RMs [20, 21].

The current view is that formal commutability studies need to be undertaken. The previous pragmatic approach was based on the not-unreasonable view that a minimally manipulated human serum sample has a higher likelihood of being commutable to clinical specimens than a delipidated, deproteinated, defibrinated plasma sample which could be spiked with compounds of unknown purity or lyophilised to enhance stability.

There is also a question on the scope of commutability testing. For example, you could be testing whether a single donation of off-the-clot serum, which has been stored frozen and thawed once and used as a single donation, is commutable with clinical specimens. Similarly, you could be testing that whether off-the-clot serum, which has been subjected to more than one freeze/thaw cycle and pooled with a number of other donations (whether defined or not) to acquire the required volume, is commutable. Alternatively, you could be testing that whether off-the-clot serum, which has been subjected to more than one freeze/thaw cycle and pooled with other donations and had individual added analytes, is commutable. The list of options goes on. These are three different scenarios which would each require commutability testing for each analyte. In terms of our definitions, only the first would be considered a 'real' material, but pooling and minimal manipulation is required to achieve the volume and range of concentrations that an EQA/PT provider is likely to require.

Concentration ranges

As a broad generalisation, many chemistry analytes have a simple Gaussian distribution in the population with low concentrations associated with a particular disease state, whilst high concentrations associated with a different disease state. The larger 'normal range' in the centre of the distribution is associated with a non-disease state. Often these concentrations are in the same orders of magnitude (*e.g.* creatinine reference range in male adults may span 70–120 µmol/L, but very high levels may only be as high as 400 µmol/L). Conversely, some hormones which are regulated in vivo by complex negative feedback loops involving both the brain and the organ of hormone production have concentrations that could cover several orders of magnitude (*e.g.* thyroid-stimulating hormone (TSH)

could be < 0.01 mIU/L in hyperthyroidism and around 100 mIU/L in hypothyroidism).

As a rule of thumb, the concentrations at the extremes tend to be the most important, but often, they are the most challenging to measure and to source suitable materials for. It is practically and ethically very challenging to obtain material from the sick or unwell patient. Taking a unit of blood, from even the most willing of volunteers who has anaemia, to allow us to have samples with subnormal levels of Vitamin B12 would just not be allowed. And rightly so, most EQA/PT providers are reliant on obtaining material that coincidentally has low/high analyte concentrations, or they have to manipulate the material either by dilution (use of inert matrix, human serum albumin, saline, phosphate buffer etc.) or by the addition of exogenous pure compound. If you are lucky, and cost and supply is not an issue—which is rarely the case in the real world—you might be able to add certified reference materials (CRMs).

For an EQA programme to be challenging, it cannot restrict itself to sending out 'beige' concentrations. To be of any value, it must be challenging. Some EQA/PT programmes are regulatory in nature and as such 'require' a large number of participants to 'pass' in order for them to continue to practice. This is often achieved at the expense of not circulating difficult-to-pass specimens, the ones where the clinical benefit is at its most rewarding. This stance is further normalised by many in the EQA/PT community calling challenging specimens 'educational' which are then excluded from performance surveillance rather than embracing these specimens as part of a value-added EQA programme. These specimens actually challenge methods and laboratories where the clinical impact is most relevant and where getting a 'wrong' answer is a very serious issue (q.v.).

The more a sample is manipulated—'diluted' or exogenous material is added—the more synthetic the material appears. Remember also that it may not be practical to add exogenous material of human origin. An example here is enzymes, where in some cases, human enzymes exist, but predominantly, enzymes from animal origin are used. This is acceptable, if the limitations of the sample preparation are taken into consideration when data are reviewed at a higher level. It may be that the EQA/PT provider is able to occasionally distribute genuine clinical material or material with very minimal manipulation at these challenging concentrations. Comparison of relative method biases between the different matrices will give confidence to the EQA/PT provider that the more routine 'highly manipulated/synthetic' approach is justified.

Volume of sample

The volume of sample that is provided to participants is always a hot topic of conversation. Participants require sufficient volume for them to undertake the analysis on all registered instruments; however, though it is poor practice, some participants like to analyse specimens in duplicate, triplicate *(unless this is their routine practice)* and have sufficient material for any troubleshooting and method verification. The use of 'real' EQA/PT testing items does usually come with volume limitations, whereas 'synthetic' EQA/PT testing items allow greater flexibility as they can be designed to meet the volume requirements. A number of EQA providers recommend that the laboratory has a registration for each individual analyser so that there is sufficient volume of material.

The number of participants in an EQA/PT programme will also impact the volume requirements. An EQA/PT provider wants to encourage as many participants as possible as this gives more power to the data that are collected. However, there may be limitations on the number of participants that can receive specimens if there are genuine volume constraints.

Volume limitations is of course related to the sample matrix. If real sweat material was to be distributed, it would not be physically possible to acquire sufficient volume from a single donor (more often a sick infant) for more than a handful of participants, certainly not the number of participants that are likely to be enrolled in the EQA/PT programme. In this case, full synthetic material is required.

A POCT EQA/PT programme may use a material from a single donor, or it may be possible to combine a number of matched bloods. The handling requirements of fresh red blood cells, to ensure sample integrity was not compromised, would still be a challenge if there were thousands of participants, and it may be necessary to have different pools/batches for different subsets of participant. The EQA/PT provider needs to ensure that homogeneity is not compromised with any material. Therefore, unless they have industrial-scale processing capacity, they may not be able to process large volumes at the same time.

Commercial organisations that provide control material to diagnostic companies or EQA/PT providers are likely to have the physical capacity and infrastructure to work with large volumes. However, when multiple donations of serum are pooled together, any interferent is likely to be diluted out. We have shown that for an analyte like creatinine, 'interfering substances' can be reduced when multiple donations are pooled together [23]. So keeping volumes small can be advantageous.

There are several advantages of having large volumes of material available. One being the possibility of sending repeat distributions to participants to test analytical performance over an extended period of time. This is really useful to demonstrate changes in analytical performance for a particular manufacturer, whether it is expected or not. Figure 4 shows the method means for two representative methods for serum folate, which have been distributed on

two or more occasions over a period of one year. Figure 4a is data for Abbott Architect, and this shows consistent mean folate concentrations across all pools, over this time period, whereas Fig. 4b, which is the Roche Cobas, shows what happened when the manufacturer actively re-formed their assay resulting in lower folate concentrations. The magnitude of the change is sufficient to move a result near cut-offs to a different clinical interpretation category. Manufacturers do change their assays, intentionally or unintentionally, and EQA/PT providers need to be able to detect this in real time. Having sufficient volume of material that can cover multiple distributions allows this.

Some EQA/PT providers will actively keep material for several years and distribute them periodically. The overhead of keeping such materials is expensive both in monetary terms and in space considerations, but the benefits of having planned such a distribution pattern are often priceless. Again, unless you have access to this type of programme design, you really cannot imagine the benefits it can bring to the value-added EQA programme.

Homogeneity and stability

Homogeneity and stability assessment of EQA/PT items are crucial, and EQA/PT providers are obliged to assess these properties as part of their ISO/IEC 17043:2023 accreditation requirements [2]. Inhomogeneity and instability can severely impact the evaluation of participant results and any conclusions drawn from the data in relation with method/manufacturer performance.

The spectrum of materials used in clinical biochemistry is wide, and different processes may be required to ensure homogeneity, depending on the material. Well-mixed, aqueous or serum-based samples are essentially simple to deal with. Whole blood material requires extra care in mixing to prevent damage to cells. Faecal materials are difficult to deal with due to their viscosity. Lyophilised material needs assessing across the racks/positions in freeze dryer to ensure no hotspots.

Homogeneity and stability assessment refer to the entire process from preparation until when a participant receives the final product in their laboratory, and indeed until the actual analysis is undertaken. This could be taken further, and stability may be required up to a period of time where a participant could ‘appeal’ the result. Obviously, this depends on the matrix of the material and the requirement of the participant to store material according to the EQA/PT providers instructions. An EQA/PT provider needs to ensure that all participants have an equal opportunity of reporting the same result, assay, and method limiting, within the allowed timescale.

An example of the challenges that an EQA/PT provider might be faced with is the faecal sample. A very common component of interest is blood haemoglobin, which is a marker of bowel cancer. Over the years, the traditional Guaiac test has been largely superseded by the Faecal Immunological Test (FIT) with its improved sensitivity. The problem comes during pre-examination. If blood were to leak into the gut lumen high up towards the stomach, then there would be more chance of the blood being subjected to churn/mixing, and less blood would be on the surface of the stool. If the blood were to leak into the gut lumen lower down, towards the rectum, then it is more likely to be present on the surface of the stool. If you were to use a ‘picker’ sampling device, then, depending on where you took your sample from, you would have a higher or lower chance of sampling from the blood-containing faeces. This is a real and genuine problem for all clinical samples, whether it is that the ‘picker’ is stabbed into a pea-sized piece of faeces or from a single point on a whole stool collection, before being loaded into a stabilising buffer in a cartridge. For EQA/PT providers, we need to avoid such errors, so mixing of material is paramount. Mixing has to occur in such a manner that the blood is not denatured, but homogeneity needs to be ensured. Even by doing this, a high degree of variation is observed which is likely to be related to the pre-analytical

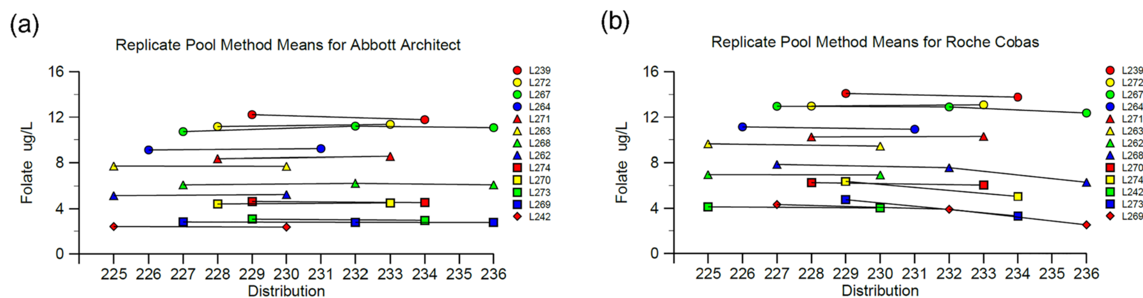

Fig. 4 Reproducibility of serum folate for a range of pools **a** Abbott Architect and **b** Roche Cobas. The identifier of each unique pool is given in the relevant legend as a pool number. The concentration of serum folate is given on the *y*-axis and time is on the *x*-axis. Each distribution represents one month. The data are from 2016

component of use of the picker. Laboratory results can be artificially improved by the provision of material direct into a buffered material, *i.e.* missing out the pre-examination step; however, this is still happening by the inexperienced patient before a specimen is presented to the laboratory [24].

There are different approaches for minimising the effect of instability. Some EQA/PT providers instruct all their participants to analyse on a particular day, storing the material 'appropriately' in the intervening period. Transit time, transit conditions, storage conditions, whether the samples were brought to room temperature before analysis, mixing, or reconstitution, can all impact the final result. The mantra—'analyse on receipt as if from a patient'—really does reduce the impact of these variables. Some EQA/PT providers use a disposable temperature tracker to log and monitor the temperature that the specimens have been stored at. This can be useful when investigating any stability-related issues. Couriers tend to record the exact details of delivery, so this is a useful source of transit data. Many EQA/PT providers undertake various behind-the-scenes calculations and make graphs although these are not necessarily shared with the participants. Nevertheless, they are used in audits and reviews to ensure that the EQA/PT data truly reflect analytical performance and not the vagaries of the postal service.

If an EQA/PT provider identifies that their material has not passed their stability and homogeneity criteria once the material has been distributed, the EQA/PT provider will have their own procedures for handling results which may include informing participants. This would be identified as non-conforming work to the EQA/PT provider and investigated accordingly.

Challenging specimens

Clinical specimens are not always routine and easy to measure. Biological systems are complex as are the spectrum of disease states. Examination systems are regularly pushed to their extremes, and the only way of knowing how the assays are performing in real-life is to test them with EQA/PT specimens. This can be achieved by constructing specimens for a specific clinical picture, for example, sending concentrations at the extreme of measurement or combining analytes at different concentrations. This is all possible with synthetic EQA/PT items, but acquisition of real material may be a lot harder. The advantage of having EQA specimens representative of a clinical picture is that it is possible to ask audit-style questions on the current laboratory practices which is then beneficial for driving change for the patient. A clinically relevant example of the type of added value the EQA programmes might deliver was from the UK NEQAS for Serum Indices programme where the investigation of low-level haemolysis in the case of hypokalaemia was probed. A serum base material was constructed with low potassium concentration. Different concentrations of haemolysate were added to achieve low levels of haemolysis. Participants were asked to measure the serum indices (including haemolysis index) and potassium. The aim of the programme is to provide EQA for the serum indices, but the by-product from this study was that it was possible to see the impact of haemolysis on potassium and also probe laboratories on how they handle specimens of this nature [25].

Although we have been describing the spectrum of specimen types, for some POCT, the sample of choice in routine testing is a tiny drop of warm capillary whole blood that you get from the patient's finger. This means that for every sample an EQA/PT provider distributes, even if it closely resembles real human blood, be it a venous collection, there is still the complete artificial scenario of trying to get it out of a tube onto the relevant device.

Interferences

Some specimens arrive in the laboratory with interferences present. These may be easily identifiable, *e.g.* through the analysis of serum indices or a drug written on a request card, or more likely, the laboratory scientist will be none the wiser to what is actually in front of them.

Good EQA/PT programmes will use their specimens to fully probe analytical systems as well as educating the laboratory and users. One of the major advantages of having a frequent distribution schedule with several samples per round is that you can fulfil the routine/standard assessment of linearity, calibration, analytical sensitivity, and so forth and have the bandwidth for some challenging specimens.

Interferents are usually added to material, so becoming a synthetic EQA/PT item, though it may also be possible to acquire material with a known interferent (*e.g.* haemolysed, the presence of rheumatoid factor antibodies confirmed). The advantages of adding an interferent is that the EQA/PT provider has control over the concentration that is added. The challenge to the laboratory is whether their assay detects the interferent. This is usually achieved by the comparison of a 'spiked' sample to a base material that has not been spiked (other than correction for any dilution effect). Examples include prednisolone interference in cortisol analysis or norethisterone in testosterone analysis. These structural homologues can cause real clinical interpretation problems. In these cases, the interference is not new, and it will be documented in the manufacturer's kit insert; however, the laboratory may not fully have appreciated how it could impact the service that they provide.

We show here an example for norethisterone [26]. Figure 5 shows pooled female serum containing different concentrations of norethisterone at concentrations that could be

found in the oral contraceptive pill. There are two key messages: (i) it is usually not just a single manufacturer's assay that has a problem; rather, it can be a whole class of methods and (ii) some manufacturer's assay change over time. Here, we have a situation where the apparent concentration of testosterone in a female patient matrix can readily increase to a level which is consistent with a number of serious disease states and which would require further investigation or treatment. The mass spectrometry methods are sufficiently selective that they have no problem in measuring the testosterone in the sample. In 2011 (Fig. 5a), fewer manufacturers of total testosterone assays were affected than in 2018 (Fig. 5b). This is directly due to manufacturers reformulating their assays to be more sensitive for testosterone at low concentrations. An important learning point from this is that assays do change over time and it is important for EQA/PT providers to continually be probing interference such as this, so that laboratories and their users are aware of the limitations of their assays and subsequently their results. This type of study would have been very difficult with 'real' EQA/PT testing items, and the limitation of adding exogenous material to serum is far outweighed by the knowledge gain on assay performance and educational stimulus for users.

Discussion

The various components of EQA/PT programme design that relate to sample production have been discussed. There is not a one-size-fits-all design for EQA/PT programmes [27]. There are some features that exemplify the unique selling points of each, and the provider will take that into consideration when designing the programme [28–30].

It cannot be argued that simply using just fit-for-purpose EQA/PT material is a sufficient requirement even if it allows a laboratory to make a pass/fail snapshot judgement on its performance. We know that data from EQA/PT programmes are used and relied upon so heavily, not only within individual laboratories but also to influence guidelines [31]. EQA/PT providers are in a unique position of having access to large number of laboratories and can readily test the current state-of-the-art, not only in terms of equipment but also in terms of analytical practices. The best EQA/PT providers are frequently invited to governmental/health department expert groups as they have a unique insight into how pathology is actually delivered, warts and all. This is so different to just relying on diagnostic companies' instructions for use claims, which even their proponents would agree are often very limited, very date, or very selective to what their legal and compliance teams can commit.

All laboratories will be measuring IQC samples on their measuring/examination systems which gives them a better understanding of any problems in real time. However, the main limitation of IQC is that by its very nature in the majority of cases, it is likely to be highly manipulated, prepared from large volumes, potentially with added preservatives. This means that it is very good at looking at the day-to-day variation of measurements, but it will tell you very little about actual assay performance and more importantly how your assay handles clinical specimens. We are often told by laboratories that there are no issues with clinical specimens or IQC specimens. How do laboratories know this? It is only by having well-designed EQA/PT programmes that a provider can challenge manufacturers or assert that there are analytical issues. The power is in the quality of the material and the amount of data that is available.

Synthetic material is often non-negotiable for some programmes because of the volumes required. This does not mean that it is second-best [32]. You can use it to your advantage to raise the bar of quality in your area of interest. Probe extended concentration ranges, construct samples that mimic clinical practice, look at cross-reactivity, perform recovery experiments to check calibration, and store material for repeat distributions over a number of years.

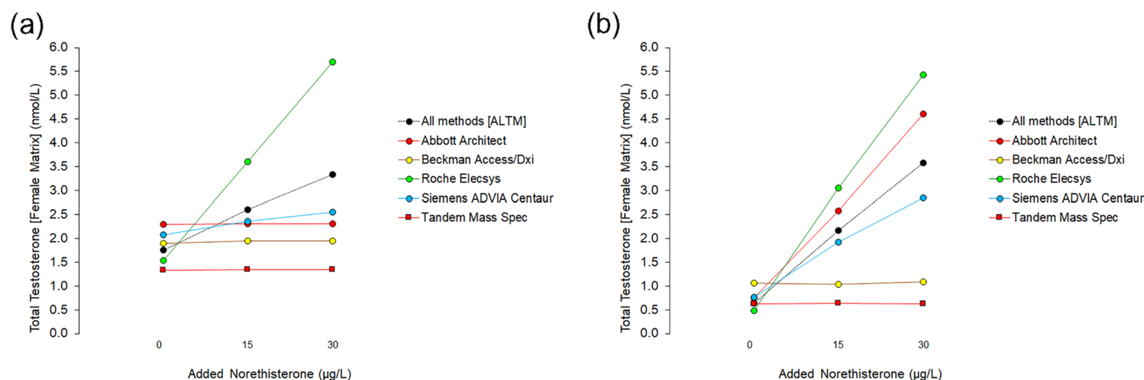

Fig. 5 Serum total testosterone results for a range of methods for a base material and increasing amounts of added norethisterone, **a** data from 2011 and **b** data from 2018

Commutability of material is important. This has been discussed in depth over recent years [33]. But it is important to realise that we are in a world where there is a wide range of EQA/PT material available. One provider's 'synthetic' EQA/PT item may be very different to that from another provider. For example, Provider A may only pool material together from the same sex donor, with minimal sample manipulation and Provider B may pool material together from any source, perform defibrination and delipidation processes, add a wide range of exogenous analytes and preservatives, and subject the material to a lyophilisation process. These are not the same and should not be classified as the same. It is probably past the point that there can be standardisation in this area, but it is incumbent on us as EQA/PT providers to educate our users about the material that we provide, the limitations of this material, and what they can get from the data.

Conclusion

The key take-home message is that in certain fields, like clinical biochemistry, there is not an outright winner between real and synthetic EQA/PT testing items. There are limitations to both. The purist would argue that only real material can be used and commutability needs to be proven, but there are clear advantages of using a mixture and a compromise approach means that deficits in one area can be countered and negated by the benefits of other materials. It would be a very dull and bland programme design that consistently distributed mid-range, normal samples, even if that was what the majority of the workload is. There must also be a balance with the expectations of the end-users who do, at times, expect material to be available for all concentrations/clinical conditions, at any volume. This is just not possible.

EQA/PT is a key component to ensure laboratory accuracy. However, along with this, it has a vital role in post-market surveillance of assays and in the education of laboratory scientists and other users of the service. To do this, it is important that we continue to send challenging specimens to laboratories and continue to drive quality improvement.

Having a mixed economy, using blood fresh out of the arm and well-designed, targeted, probing synthetic material where the particular situation demands it, is a pragmatic but sensible approach in our opinion.

Author contributions All authors reviewed the data, and wrote and reviewed the main manuscript. FM prepared the figures.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article. The authors did not receive support from any organisation for the submitted work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Sciacovelli L et al (2010) The role of the external quality assessment. Biochemia Medica 20(2):160–164
2. ISO, (2023) ISO/IEC 17043:2023 Conformity assessment: general requirements for the competence of proficiency testing providers. ISO. p36
3. ISO, (2022) ISO 15189:2022 Medical Laboratories: Requirements for quality and competence. p62
4. ISO, (2017) ISO/IEC 17025:2017 General requirements for the competence testing and calibration laboratories. p30
5. Lee J-S et al (2017) Effects of pyridoxal-5′-phosphate on aminotransferase activity assay. Lab Med Online 7(3):128–134
6. Birmingham Quality, (2023) UK NEQAS for clinical chemistry distribution 1160 commentary. [accessed 19th February 2024]
7. Alsaeedi A, Welham S, Rose P (2023) Impact of lifestyle factors on dietary vitamin B(6) intake and plasma pyridoxal 5'-phosphate level in UK adults: national diet and nutrition survey rolling programme (NDNS) (2008–2017). Br J Nutr 130(8):1403–1415
8. Birmingham Quality, (2022) UK NEQAS for steroid hormones distribution 499. [accessed 19th February 2024]
9. Sarigul N, Korkmaz F, Kurultak İ (2019) A New Artificial Urine Protocol to Better Imitate Human Urine. Sci Rep 9(1):20159
10. Sri, A., et al., (2015) Preparation of artificial saliva formulation. International conference icb pharma II
11. Harvey CJ, LeBouf RF, Stefaniak AB (2010) Formulation and stability of a novel artificial human sweat under conditions of storage and use. Toxicol In Vitro 24(6):1790–1796
12. Zivkovic AM et al (2009) Effects of sample handling and storage on quantitative lipid analysis in human serum. Metabolomics 5(4):507–516
13. Cramb R et al (2008) Lipid external quality assessment: commutability between external quality assessment and clinical specimens. Ann Clin Biochem 45(Pt 3):260–265
14. Miller WG et al (2018) IFCC working group recommendations for assessing commutability part 1: general experimental design. Clin Chem 64(3):447–454
15. Larsson M, Nilsson G (2023) Consequences of insufficient selectivity in quantitative and qualitative chemical analysis. J Anal Sci Method Instrum 13:13–25
16. Delanghe JR, Speeckaert MM (2011) Creatinine determination according to Jaffe-what does it stand for? NDT Plus 4(2):83–86
17. Marrington R, MacKenzie F (2023) Variation of eGFR reporting and CKD equations used in the United Kingdom. Ann Clin Biochem 60(5):328–338
18. Greenberg N et al (2012) Specificity characteristics of 7 commercial creatinine measurement procedures by enzymatic and Jaffe method principles. Clin Chem 58(2):391–401
19. Nilsson G et al (2018) IFCC working group recommendations for assessing commutability part 2: using the difference in bias

between a reference material and clinical samples. Clin Chem 64(3):455–464
20. Budd JR et al (2018) IFCC working group recommendations for assessing commutability part 3: using the calibration effectiveness of a reference material. Clin Chem 64(3):465–474
21. Miller WG et al (2020) 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 66(6):769–778
22. Sandberg S et al (2023) Recommendations for setting a criterion and assessing commutability of sample materials used in external quality assessment/proficiency testing schemes. Clin Chem 69(11):1227–1237
23. Marrington R, MacKenzie F (2024) Understanding the limitations of your assay using EQA data with Serum Creatinine as an example. Clin Chem Lab Med. Accessed 12 Mar 2024
24. O'Driscoll S et al (2020) An evaluation of ten external quality assurance scheme (EQAS) materials for the faecal immunochemical test (FIT) for haemoglobin. Clin Chem Lab Med 59(2):307–313
25. Birmingham Quality, (2023) UK NEQAS for serum indices distribution 177 commentary. [accessed 19th February 2024]
26. Jeffery J et al (2014) Norethisterone interference in testosterone assays. Ann Clin Biochem 51(Pt 2):284–288
27. Miller WG et al (2011) Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 57(12):1670–1680
28. Jones GRD et al (2017) Analytical performance specifications for external quality assessment—definitions and descriptions. Clin Chem Lab Med 55(7):949–955
29. Ricós C et al (2022) External quality control in laboratory medicine. Prog future Adv Lab Med 3(3):221–242
30. Blasutig IM et al (2023) External quality assessment practices in medical laboratories: an IFCC global survey of member societies. Clin Chem Lab Med 61(8):1404–1410
31. NICE, (2023) Jaundice in newborn babies under 28 days Clinical guideline [CG98]. Last updated 31 October 2023. https://www.nice.org.uk/guidance/cg98. Accessed 19 Mar 2024
32. Wood WG (2006) Problems with the external quality assessment of accuracy of point of care devices (POCD) for blood glucose are independent of sample composition. Clin Lab 52(7–8):345–351
33. Badrick T et al (2022) Interpreting EQA-understanding why commutability of materials matters. Clin Chem 68(4):494–500

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.