

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Practical Laboratory Medicine

journal homepage: www.elsevier.com/locate/plabm

Biological variation: Understanding why it is so important?

Tony Badrick

Royal College of Pathologists of Australasia Quality Assurance Programs, St Leonards Sydney, NSW, 2065, Australia



ARTICLE INFO

Keywords:

Analytical goals
Biological passport
Patient based quality assurance

ABSTRACT

This Review will describe the increasing importance of the concepts of biological variation to clinical chemists. The idea of comparison to 'reference' is fundamental in measurement. For the biological measurands, that reference is the relevant patient population, a clinical decision point based on a trial or an individual patient's previous results. The idea of using biological variation to set quality goals was then realised for setting Quality Control (QC) and External Quality Assurance (EQA) limits. The current phase of BV integration into practice is using Patient-Based Real-Time Quality Control (PBRTQC) and Patient Based Quality Assurance (PBQA) to detect a change in assay performance.

The challenge of personalised medicine is to determine an individual reference interval. The Athletes Biological Passport may provide the solution.

1. Introduction

Measurement is defined by the International Bureau of Weights and Measures [1] to be the process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity. The measurement process commonly used in laboratory medicine uses a ratio scale, where proportions are constant, and there are no zero points [2]. Other ratios, particularly the 'signal to noise,' will play a part in the discussion.

To interpret the result of a clinical laboratory test requires comparison to a reference interval, a clinical decision point, or previous results. A reference interval is derived from a *reference* population, whereas detecting a *significant* change in consecutive results requires an understanding of analytical and biological variation (BV). Even a clinical decision point requires the ability to separate two patient populations. The common theme running through this article is biological and analytical variation.

The reason for using BV as a component of performance goals is that to detect or rule out disease, an assay must be able to see a change from normal. Thus BV, or specifically, being able to detect a difference in a measurand 'signal to noise' of BV to analytical precision is the basis for many goals, including:

- how quality specifications are set often linked to the sigma metric [3,4],
- method selection and assessment [5],
- the definition of common reference intervals in a geographic area [6,7],
- the analysis of the consequences of poor calibration [8],
- evaluation of laboratory results in external quality assurance programs (EQA) [9],
- the quality of evidence-based clinical guidelines [10], and,
- goals for PoCT performance [11].

E-mail address: Tony.badrick@rcpaqap.com.au.

<https://doi.org/10.1016/j.plabm.2020.e00199>

Received 25 October 2020; Accepted 23 December 2020

2352-5517/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The principles behind setting performance goals look at the impact of bias and precision on patients' misclassification. Analytical bias or precision will move a patient result relative to a reference interval or clinical decision point. Petersen has reviewed these situations [2,12]. There are two general models for classifying a result in a patient based on a reference interval, either a unimodal or bimodal. In the bimodal approach, there is a decision point based on the prevalence of the disease. There is an assumed increased risk in the unimodal modal as the measured concentration approaches a decision point or reference limit. The decision point/reference limit separates the otherwise homogeneous distribution into two groups with high and low risk of some disease, respectively. This approach contrasts with the bimodal distribution in which there is a well-defined (but often unknown) prevalence for the condition under consideration. Typically, the assay's performance in identifying the diseased group is described using the fraction of true positives, true negatives, false positives, and false negatives. In the bimodal model, the impact of bias and precision is dependent on the prevalence. Petersen et al. have also investigated the effect of bias and precision on ordinal measurands such as dipstick tests [13]. However, there is often an interplay between data-based and theory-based models as a deeper understanding of these problems' complex nature. A model is posed and used until it fails some significant observation. A data-based approach is used until a newer model can be developed based on the fresh insights provided by more data mining.

There have been some recent reviews of Biological Variation (Callum G. Fraser, 2017; Ricos et al., 2009) as well as an excellent reference text [16] and Chapter [17].

This Review aims to track the history of how biological variation has influenced the Total Error concept, Analytical Goal setting, Quality Control (QC), EQA practice, through Patient-based real-time Quality Control and the future of EQA using patient population parameters.

2. Biological Variation

BV for a biomarker entails in each individual a "subject mean" or central tendency, control level, or "setpoint" concentration of homeostatic regulation arising from such factors as genetic characteristics, diet, physical activity, and age [18,19].

BV's formal study began in 1835 though understanding differences between individual humans or animals is a fascination and necessary skill for any human. This formal study was driven by the development of statistical tools created in the scientists of the age of Enlightenment (Fermat, Pascal, Newton, Brahe, Galileo, Daniel Bernoulli, Laplace, Gauss). The equation describing the Gaussian distribution came from investigating errors in astronomical measurements published in the early eighteenth century by Gauss [20]. The first significant application of the Gaussian distribution to humans' measurement was by an astronomer turned social scientist named Adolphe Quetelet [21,22]. His most influential book was entitled *Treatise on Man*, but a literal translation of the French would be "On Man and the Development of his Faculties, or Essays on Social Physics." Quetelet described the concept of the *average man* who has variables, encompassing social and physical characteristics that follow a normal distribution. Quetelet believed that understanding the statistical description of these variables would reveal God's work as well as being a force for good administration. Using this approach, he gained insight into the relationships between crime and other social factors, including age, gender, climate, poverty, education, and alcohol consumption. Quetelet postulated a theory of human variance around the *average*, with human traits being distributed according to a 'normal' curve. He proposed that normal variation provided a basis for the idea that populations produce sufficient variation for artificial or natural selection to operate.

This concept of a 'normal range' influenced laboratory medicine as it was possible to classify 'normal' from 'abnormal or diseased.' However, the question of what is 'normal' or 'average' arises. Consider a disease such as sickle cell anaemia. It is a disease, yet it is protective against malaria, so the abnormal in one population becomes normal. The diseased in one group represents non-diseased in another. What is normal or average in one population is not what is healthy.

What is the 'normal' population? Note that we use the words 'normal' and 'Gaussian' almost interchangeably. Poincare pointed out that mathematicians believe the normal distribution is an empirical fact and that to scientists, it is a mathematical law [23]. This view has not changed to this day. This is an example of a theoretical model we sometimes use despite data that indicates the contrary.

To explore this further, some interpretations of the concept 'normal' are given below in Table 1 by Murphy.

3. Reference intervals

Laboratory medicine started developing a broader range of tests during the mid-twentieth century. These tests were developed inhouse and provided to a local medical community, usually in isolation from other sites. At this time, the clinical practice was to compare a patient's results with an ill-defined or at least inconsistently defined, range of values or 'normal limits,' called the 'normal

Table 1
Interpretations of 'normal' (modified from Murphy [24]).

	Conceptions of normal	Suggested alternatives
1	Determined statistically	Gaussian
2	Most representative of its class	Average, median, mode
3	Most commonly encountered	Habitual
4	Wild type: most suited to survival & reproduction	Fittest
5	Harmless 'carrying no penalty.'	Innocuous/harmless
6	Most often aspired to	Conventional
7	The most perfect of its class	Ideal

range.’ This was derived from a population of supposedly ‘normal’ (meaning healthy) individuals [25]. It soon became apparent that multiple ‘normal ranges’ were required for different patient populations and individual laboratories, and that these were different because of methodological variation. It was the pioneering work of Sunderman who revealed using the first EQA surveys the extent of the variation in results caused by these different in-house methods [26].

Grasbeck and Saris in 1969 [27] challenged this concept of a normal range, arguing that normality is a relative term and situational. They suggested a new name, the ‘Reference Interval,’ which would describe fluctuations of analyte concentrations in well-characterised groups of individuals. The fundamental idea was to have a point of reference against which to interpret an individual’s results, rather than defining normality [28]. The population reference interval may not account for factors such as age, ethnicity, or gender unless they have a significant impact. Therefore, the reference interval approximates what can be expected in the population from which the patient belongs.

3.1. Components of BV

Many of the analytes we measure in laboratory medicine change over time, varying in predictable rhythms; indeed, sometimes the loss of these rhythm represents disease. However, most quantities do not have these rhythms, and for each individual, the quantity varies around a homeostatic set point in a truly random manner [29]. This is the true meaning of BV, though the variation we see will be due to different setpoints and variation for an individual, often lost in the total variation. Of course, the added variation is caused by the pre-analytical processes of collection, transport, preservation, and the measurement system’s variation.

The underlying reasons for using biological variation in the many situations described in this article are compelling; however, the value relies upon accurate estimates of that variation. Many investigators have produced estimates for the intra-individual (CV_I) and group (CV_G) variation, but often the needed measurement rigor and supporting information was not available [15,30–34]. For example, there have been studies conducted with small sample sizes and poorly defined groups with inappropriate statistical methods. There is now a better understanding of the need to produce and promulgate accurate estimates generated from significant sample sizes using the best statistical tools available [30]. These biological variation estimates are available at the curated EFLM site [30]. Important statistical considerations include determining the BV parameters, outlier removal, and their confidence intervals.

The method of determining CV_I and CV_G involves selecting a group of reference individuals in *steady-state* and measuring the analyte of interest at regular intervals over a period which will reflect any ‘normal’ biological variation. The variation will, of course, also involve any additional variation due to the collection and measurement process. The analytical variation is estimated by measuring replicates of the individual samples. The statistical analysis used is ANOVA, which can identify the significant variations due to intra-individual, group, and analytical causes. The specific ANOVA model is a nested random design in two levels, assuming that the three terms are independent and normally distributed with constant variances [14,35]. Historically the sample sizes have been small of the order of 10–20 individuals. However, Roraas et al. [36] have produced power tables with numbers of individuals, samples, and replicate measurements providing guidance. They also offered a ‘width’ of the confidence interval based on the numbers above to understand the limits’ uncertainty. There is skewness around the confidence intervals depending on the ratio of analytical and intra-individual biological variation, another example of the impact of noise on measurement.

3.2. Index of individuality

As mentioned earlier, it was only in 1956 that Williams first established intra-individual biochemical variation based on published studies of various biomarkers [29]. He hypothesized that the ‘normal’ man is an illusion and that every individual has some variation in some respect. Further, he postulated the ‘principle of genetic gradients,’ which stated that, ‘Whenever an extreme genetic character appears in an individual organism, it should be taken as an indication (unless there is proof to the contrary) that less extreme and graduated genetic characters of the same sort exist in other individual organisms.’ Williams termed this ‘biochemical individuality.’ Some analytes are tightly controlled within an individual around a setpoint, but the setpoint may vary between individuals; the classic example is creatinine. The significance of this to diagnosis is that an individual may be within a reference interval for the population but have significant renal function loss. With these analytes, there is a need to have more granularity in the reference interval by partitioning by age, sex, and body mass; however, this may still not be sufficient. A comparison with previous results may be necessary. However, there have not been many databases containing histories of patient results that could be used for this purpose in practice until recently.

Recognising that with some analytes there is significant intra-individual variation, the index of individuality was introduced by Harris [37] as the ratio (combined biological within-subject variation and analytical precision divided by the biological between-subject variation), but is now often simplified to $SD_{intra-individual}/SD_{group}$. The use of the index allowed the distinction between situations where the population-based reference intervals were relevant (index > 1.4), and those (index < 0.6) for which use of cumulative (with respect to time) systems for reporting results for individuals would be favoured. In those situations, the population should be stratified into more homogeneous subgroups (each of which, of course, would have a higher index). Still, ideally, use should be made of individual reference intervals.

Note that the index of individuality is another example of using the signal to noise ratio to determine when a measure is useful. It also demonstrates that more data challenges assumptions on models and leading to improved understanding and new models.

3.3. Biological Passport

Personalised medicine (PM) has the potential to tailor therapy with the best response and highest safety margin to ensure better patient care [38]. In Laboratory Medicine, the concept would involve personalised reference intervals [39,40]. These ranges have been considered for over a decade or so but are technically challenging to generate in practice [41,42].

However, the sports anti-doping agencies developed an Athlete Biological Passport (ABP) to identify variations in an individual's biomarker profile, which promises to achieve the goal of a model suitable for a personalised reference interval. The ABP uses Bayesian statistics as its underlying model and is based on previous success with clinical trials [43]. Bayesian analysis uses prior knowledge of the inter-individual variation of a biomarker measurement to progressively integrate previous results and represent the levels more accurately for an individual. Each new measurement of the biomarker is compared to a critical range, which is derived from the population reference interval. Then, this range progressively adapts with each unique measurement point. Bayes theorem predicts that the probability of an effect given each cause is the likelihood ratio [44–46].

Three distinct modules can be distinguished in the Biological Passport: the hematological, steroidal, and endocrinological modules. The hematological module of the passport aims to detect any form of blood doping. In 2008, the Union Cycliste Internationale was the first sports organization to implement the hematological module of the passport to deter blood doping in elite cycling. The steroidal module aims to detect direct and indirect forms of doping with anabolic agents [47–50]. The endocrinological module measures biomarkers such as IGF-1 and growth hormone. There are further modules being investigated, such as the 'omics' module [51,52].

The model can be complicated and considers specific genetic variation to the metabolism of an analyte. Thus, over a reasonably short period, an individualised reference interval is created. Any significant variation is then identified, which may represent doping or an individual's case, a variation towards disease [44]. It is possible that using this type of logic will lead to a workable model of individual reference intervals that will be created and various routine measurements that could feed into the model to detect incipient disease. The inputs will eventually come from wearable devices [53,54].

4. Analytical goals

Any measurement made needs to be fit for the purpose for which it is to be used. In laboratory medicine, that purpose is to differentiate between 'healthy' and diseased. In the previous section, we have discussed the problems associated with defining what is health. Next, we will investigate how this can be achieved with our measurement processes. We will limit ourselves to measurement procedures that give numerical results on a ratio scale, so the measurement parameters include accuracy, precision, reproducibility, and repeatability. The required accuracy and precision of a measurement are termed the quality goal for that analysis.

The importance of defining quality goals has been a continuing global effort, and in 2014 in Milan, there was an update to the Stockholm 'Consensus Agreement' [55,56]. Revising the original goals involved modification and explanatory additions to simplify the hierarchy and improve its application by various stakeholders.

This simplified scheme resulted in three models, as shown below:

Model 1. Based on the effect of analytical performance on clinical outcomes

This can, in principle, be done using different types of studies:

- 1a Direct outcome studies – investigating the impact of analytical performance of the test on clinical outcomes. This approach requires a Randomised Clinical Trial, which is impractical for reasons of non-uniqueness of the biomarker/disease except in a few conditions where there is a well-defined clinical pathway with close links to outcomes – e.g., HbA1c, Troponin, cholesterol [57, 58].

These tests have decision limits (established based on existing analytical performance) and are used as a part of disease definition, or they are used as proxy outcome measures to assess patient well-being or response to treatment.

- 1b Indirect outcome studies – investigating the impact of the test's analytical performance on clinical classifications or decisions and thereby on the probability of patient outcomes, e.g., by simulation or decision analysis. This model is suitable when the clinical findings associated with the test results are well defined. There is evidence about the diagnostic accuracy of the test that is generalizable to the patient population setting, and the consequences of correct/incorrect classification have already been established.

Examples of goals based on clinical outcomes using modeling or simulation include the following; fasting plasma glucose (FPG) measurement on the misclassification of Diabetes Mellitus (DM) [59] and IFG [60]; FPG, OGTT, HbA1c measurement on classification of DM [61]; point of care glucose meter results on potential insulin dosing errors [62]; direct measurements of HDL and LDL cholesterol affected by bias due to hypertriglyceridaemia on cardiovascular risk prediction [63]; and, cardiac troponin methods on the misclassification of patients presenting with Acute Coronary Syndrome to Emergency Departments [64,65]. An example of setting performance goals based on a survey of clinicians was described by Panteghini [66].

Model 2. Based on components of biological variation of the measurand

This model has the advantage that it can be applied to most measurands for which group or intra-individual biological variation data can be established.

Model 3. Based on state-of-the-art

State-of-the-art represents the highest analytical performance level achievable or achieved by a certain percentage of laboratories in an EQA scheme.

The highest goals are based on model 1, clinical outcome, which should be specifications based on decision levels and evaluation of tolerable false-positive (and false-negative) results. The problems with this approach include there are few tests have a definitive role in managing a specific.

Condition's clinical pathway and most laboratory tests are used for both diagnosis and monitoring disease, often in combination with other laboratory or medical tests. Thus the link between testing and health outcomes is indirect [67].

Thus, it is unlikely that clinical outcomes will be suitable to define performance goals for every measurand in every clinical situation. Therefore, the next best approach will be to use model 2, quality specifications derived from biological variation data. The principle behind this approach is to minimize the "signal-to-noise" ratio between analytical variation and the (natural) biological variation. Thus, analytical performance specifications based on biological variation are now broadly used in clinical chemistry, whereas in clinical guidelines, performance based on clinical criteria is, in most cases, preferred [68,69].

Setting performance goals to improve analytical performance may not improve patient outcomes [70]. The best performing assay does not lead to clinical action or affect patient compliance with treatment [71].

These specifications give the maximal total allowable variation in bias and precision, or the Total Allowable Error (TE), that will be tolerated before the assay will no longer be fit for the clinical purpose of identification of abnormal. We will discuss this concept next.

4.1. Total (allowable) error

Total Allowable Error began in 1974 when Westgard et al. [72] described a model for formulating criteria that can be used to judge whether an analytical method has acceptable precision and accuracy. The 'total analytic error' was defined as the sum of random analytic error (RE) and systematic analytic error (SE), which is, in turn, the sum of proportional (PE) and constant error (CE). For each of these errors (RE, PE, and CE), and Total Error (TE = RE + CE + PE = RE + SE), these authors devised an acceptable level of error (Table 2). This paper also describes how each of these components is calculated. An acceptable analytic error was defined to be 95% of acceptable medical error (E_A). The acceptable medical error was not determined but seen to be the limiting value of the TE equation.

The model for the equation describing TE, however, became the linear sum of bias and stable precision (s), TE = B + z*s, the z being the z-value corresponding to a defined one-sided confidence probability. This assumes that the TE is composed of maximum values of its components B, z, and s; there is a linear relationship between B and s [73,74]. However, this is not always the case. This equation's significance is that it combined RE and SE for the first time, hence 'total'. The components SE and RE are determined from linear regression equations, a concept that flows into the TE equation structure, which is the sum of two parts.

One assumption is that the bias is unchanged over time, 'Systematic' implies a specific point in time. The bias here is estimated from the linear regression based on recovery experiments. The concept of bias was a relative one, bias from and existing method that may or may not be a reference method. As Klee [75] pointed out, bias depends on the time interval considered. Although bias should be removed when possible, bias is inevitably encountered in some circumstances, such as systematic differences between analyzers measuring the same analyte.

Related to TE is the concept of an Error Budget [73,76], which has systematic and random error components shared within the analytical goal, i.e., within the specified tolerance between true concentration X_C and reported result X_R (Fig. 1) [77]. The sum of the errors must be less than the TE limit.

Krouwer [78] showed that the original approach for estimating TE using just precision and bias can give a misleading impression of assay performance because estimates of error from various individual sources (e.g., drift, bias) are generally not combined to give a picture of total error (i.e., the error that might be observed by a clinician). He suggested various models to estimate total analytical error, but these are impractical for most laboratories because of the resources needed to identify all his variables. An additional complication is that Krouwer's total error models require quantification of nonspecific interferences.

Table 2
Performance criteria for analytic errors.

Analytical Error	Experiment	Acceptable Error
RE	Replicates	$2 * SD_{TU} < E_A$
PE	Recovery	$ \%R_{(U \text{ or } L)} - 100 _{X_C} < E_A$
CE	Interference	$ Bias + t * (SD_d / \sqrt{N}) < E_A$
SE	Patient Comparison	$ (a + bX_C \pm W) - X_C _C < E_A$
TE	Replicates and Comparison	$ (a + bX_C) - X_C + \sqrt{\{(SD_{TU})^2 + W^2\}} < E_A$

Abbreviations used in the Table. E_A Acceptable medical error; X_C concentration at critical decision level; SD_{TU} analytical SD at upper 95% level; %R_(U or L) %recovery at upper or lower recovery experiment; t is t-test level at p = 0.05; W is the width of the confidence band; a, b is the intercept and slope from the regression of critical test concentration.

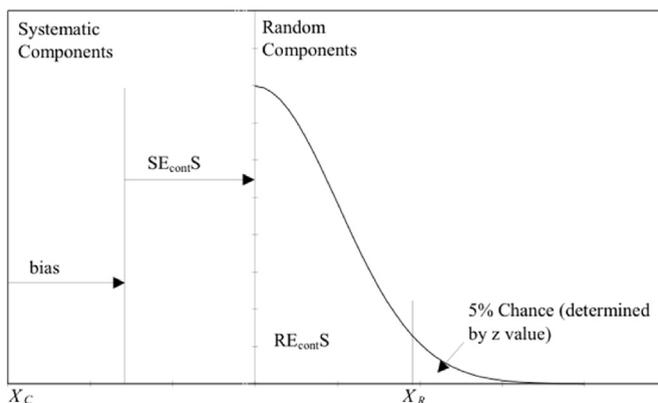


Fig. 1. Representation of standard error budget components within an Analytical Goal.

Next, we will analyse the various assumptions in determining these two major components or the TE equation’s bias and precision.

4.2. Setting quality specifications for laboratory medicine based on biological variation

Determining Quality specifications using the BV components of within and between-subject variation, have been proposed by various professional groups [79–82]. Specifically, BV’s appearance in many laboratory medicine areas was driven by the Stockholm Consensus Conference and the Milan meetings, as described earlier. The Stockholm meeting produced a unified approach to a hierarchy of performance specifications, which was generally accepted by the laboratory medicine community, even though these different approaches had been discussed for many years [83–85].

These meetings proposed that BV be used above ‘state of the art’ and professional opinion when setting performance specifications. Only data from clinical outcomes studies were superior when setting performance specifications. Note that the ‘usual’ source of ‘state of the art’ information was EQA programs. So EQA data was used to set performance goals, yet for most EQA programs, the ‘target’ values were not traceable. The realisation that performance specifications can be unified using BV for many different areas of measurement, including Reference Intervals and Clinical Decision Points, Method Evaluation, QC, and EQA performance limits then followed [56, 86–88].

There were always simplifications; however, even though most biological distributions are log-Gaussian or non-Gaussian, a simple Gaussian distribution was used. However, this assumption has been challenged. The effect of different distributions on the calculation of intra-individual BV (CV_I) has been investigated and shown to have a significant impact [36].

We will now deal with the derivation of the BV components of the TE equation.

4.3. Quality specification for analytical precision

The idea of using a biological variation to set a limit on acceptable analytical precision was due in no small part to Tonks [89]. Using observations of the differences between 170 participating laboratories in a Canadian EQA scheme, he formulated their performance goals. empirically, Tonks stated that the allowable error, which was taken to be twice the CV_a, should be:

$$2 * CV_a = 0.25 * (Reference\ mean / Mean)$$

Barnett [90] had suggested that desirable precision should be such that any analytical error did not significantly widen the normal range. This will be achieved he demonstrated if the S_A (analytical SD) does not exceed one-twelfth to one-twentieth of the normal range. If the S_A is one-twelfth of the normal range, it will widen that range by 5.4%, whereas if it is one-twentieth of the normal range, it will widen that range by 2%. These numbers were arbitrary but were based on the impact of misclassification of a patient’s result relative to a reference interval [8,91].

Cotlove et al. [18] described quality specifications for precision based on the idea that analytical precision should not affect the interpretation of results used to monitor patients rather than diagnosis, as used by Barnett. By looking at the impact of an analytical error in interpreting serial results in a patient, this group set performance goals using intra-individual variance (S_I²) with the assumption that the analytical error influence would not change between measurements. They proposed that to minimize any artifactual increase in total variation S_T = √(S_I² + S_A²), the S_A = 0.5*S_I. They showed that above this level S_T increases by 40% or more if S_A ≥ S_I.

Cotlove et al. utilised what became the Reference Change Value (RCV) or Critical Difference concept [92]. The RCV is defined to be what would be a significant change in results in a patient over two results, assuming there is no change in the analytical precision (or bias):

$$RCV = z * \sqrt{(S_I^2 + S_I^2)} = z * \sqrt{2 * S_I^2} \tag{1}$$

If there is a change in the analytical system, we can add the two analytical errors in this equation: precision or bias. Adding analytical error S_A and bias B , too (1) yields:

$$RCV = z \cdot \sqrt{(S_I^2 + S_A^2)} \quad (2)$$

Usually, the assumption is that there is no significant drift between the same analyte measurements on the same patient, but this may not be valid, mainly if the laboratory involved uses multiple instruments to measure the same measurand. There are some forms of patient-based quality assurance techniques used, such as delta checks, which seek to identify misidentification by comparing previous to the most recent results where a 'significant' difference is greater than the RCV [93]. There have been many variations on the concept, but the value of this has been challenged because of the low detection sensitivity and widespread use of barcodes [94–97].

Harris [37] investigated the impact of analytical precision on the misclassification of a patient using a reference interval, clinical decision point, or in the case of monitoring. He found that a decrease in the ratio of analytical precision to biological variation (inverse of normal signal to noise) of up to 50% leads to only a four percent improvement in overall variation (S_T). This change in analytical precision produced only minor changes in clinical sensitivity and specificity. His investigation of the impact of analytical precision on monitoring a single patient also showed a limited influence with the CV_A/CV_B of <0.5 . These findings provided theoretical support to the Barnett criterion [37,85].

Thus, the goal of analytical precision was set and validated. 'A striking feature is the fact that all of the individual approaches recommend numbers for analytical standard deviation near or equal to 0.5 times the biological standard deviation' [98].

4.4. Allowable bias

The introduction of analytical bias can lead to an entire community's misclassification as the population of results moves relative to a fixed reference interval or decision point [99]. Harris [85] expanded his earlier work to include the impact of bias by simply having a bias (B) in the goal, $S_A \leq 0.5 \cdot S_I$ became $\{S_A^2 + B^2\} \leq 0.5 \cdot S_I$.

The approach to dealing with bias involves quantifying patients' misclassification caused by the shift in patient results relative to the reference limits. This may be an unrecognised problem. When a clinical decision point is used, the laboratory must be using a method/calibration which is traceable to the laboratory where the decision point was determined, which is not often the case [100].

Gowans et al. [101] used the reference interval (Gaussian) model generated using the guidelines of the Expert Panel on Theory of Reference Values of the International Federation for Clinical Chemistry [102] as a starting point. Then, using known sample size, they calculated the effect of different confidence intervals ($z = 0.8, 0.9, 0.95$ or 0.99) and sample sizes (10–10,000) on the percentage of the population outside mean $\pm 1.96 \cdot SD$, i.e. the reference limits. Incidentally, they found that above a sample size of 800, the confidence interval for the reference limits becomes negligible.

This confidence interval for 120 patients around each reference limit (upper and lower) becomes $\pm 0.25 \cdot S_B$, where S_B is the total biological standard deviation ($S_B = \sqrt{\{S_I^2 + S_G^2\}}$), which can then converted into a maximal allowable bias subject with a 90% confidence interval. The maximum acceptable percentage of the population outside the limit for the 0.90 confidence interval of each of the mean $\pm 1.96 \cdot SD$ reference limits is 4.6% for a population sample size of 120. Based on this, the maximum acceptable precision, with no bias, is 0.6 of the total biological standard deviation, and the maximum acceptable bias, with no precision, is $0.25 \cdot S_B$.

This leads to 1.3 and 4.4%, respectively, of the population outside each (lower and upper respectively) reference limit. This was defined as an acceptable misclassification error, and hence any laboratory with a method bias less than this value could use the reference interval from the reference laboratory/method.

Stamm [103], using a similar approach to Gowans, proposed that BV's ratio to analytical between-day precision should be ≥ 2 . If the BV is unknown, then the quotient Reference Interval/Analytical between-day precision should be ≥ 8 .

4.5. Combining bias and precision

The assumption so far has been to see the two components of error as not being related. Still, precision will also lead to patients' misclassification as the precision 'adds' to the total variance and widens the 'apparent' reference interval.

We assume that the same percentage of misclassified patients is acceptable and a linear relationship between the two errors. We can calculate the maximum bias and precision that will produce 4.4% of the population lying outside the reference limits. These values are $0.275 \cdot CV_B$ for bias (when the precision is zero) and $0.597 \cdot CV_B$ for the precision (when the bias is zero). Thus, TE in this model, is not a constant but varies from $0.275 \cdot CV_B$ to $z \cdot 0.597 \cdot CV_B$. With 4.4% of the new upper limit of the distribution, the equivalent z-value becomes 1.68.

The EGE Lab Working Group combined these two errors, Cotlove's requirement for maximum allowable precision with Gowan's maximum acceptable bias into a two-component linear model [104]. This was the worst case. In a broad review of analytical goals, Petersen and Horder [105] reviewed the influence of changing bias and precision on reference intervals and decision limits. The signal underpinned their approach to noise concept where the noise could be either controllable (diet, posture, biological cycles) or non-controllable (intra-individual variation) and they also combined the maximum bias and precision into a goal. The goal becomes;

$$TE \leq | \pm \{1.65 \cdot (0.5 \cdot CV_I) + 0.25 \cdot \sqrt{(CV_I^2 + CV_G^2)} \} | \quad (3)$$

The next step was to consider the inter-relationships between these two terms. With the same assumptions as Gowans et al. we have that [106]:

$$B = (1.96 + 0.25) * \sqrt{(CV_I^2 + CV_G^2)} - 1.96 * \sqrt{(CV_I^2 + CV_G^2 + CV_A^2)} \quad (4)$$

The second component represents the dispersion of the population-based reference interval.

Petersen et al. [107] compared the different approaches to combining the bias and precision concepts described above. Using a transformation to the axis B/CV_B versus CV_A/CV_B revealed the relationship between bias, precision, and BV more closely. Gowans' specifications for sharing common reference intervals become a cusp with a maximum at CV_A/CV_B of 0.5. The EGE Lab working group's model becomes a rectangular function when plotted on B/CV_B versus CV_A/CV_B . This plot did emphasize the impact of the ratio of CV_I/CV_B , which is the index of individuality described earlier. If the $CV_I = 0.7 * CV_B$, then as an example, with a bias of zero, the desirable level for CV_A/CV_B is twice the level of when $CV_I = 0.3 * CV_B$.

Note the use of both CV and SD in these calculations. CV allows the move to dimensionless formulae but does assume that the SD increases linearly with concentration to produce a constant CV [108]. Skendzel et al. [109] described medically useful analytical coefficients of variation (CV_A) were calculated in the following way for monitoring: $CV_A = \text{medically significant difference} / 1.65 * \sqrt{2}$. Also, Fraser et al. [79] described a general theory for the allowable analytical precision when patient monitoring is involved. They found that $CV_A \leq \sqrt{\{[\Delta^2 / 2 * z^2] - CV_I^2\}}$ where Δ is the significant clinical change as a percent.

4.6. Quality specification for external quality assurance

The Stockholm consensus's impact was that the EQA Organizers Working Group [98] looked at applying the TE concepts to EQA performance specifications. Their goal was to ensure that analytical performance did not influence clinical strategies, so suggestions for acceptable limits involved understanding the purpose for which the measurand was to be used. In the monitoring situation, $S_A \leq 0.5 * S_I$ (in the absence of unidirectional systematic changes), or $\Delta SE \leq 0.33 * S_I$ (when precision is negligible). For analytes that would be used for diagnostic testing $B \leq 0.25 * S_B$ (when precision is insignificant) and $S_A \leq 0.58 * S_B$ (when bias is minor). However, as we saw earlier, the same analyte can be used for different purposes in different clinical situations.

Fraser and Petersen [106] further proposed that if there was only a single determination of each EQA sample, the 95% acceptance range for each laboratory from the target value should be the sum of both values:

$$95\% \text{ acceptance range} = \text{target value} \pm [1.65 * (0.5 CV_I) + 0.25 * \sqrt{(CV_I^2 + CV_G^2)}]$$

$$TE = 1.65 * (0.5 * CV_I) + 0.25 * \sqrt{(CV_I^2 + CV_G^2)} \quad (5)$$

Although the purpose of expression (5) was the application in EQA, it is used in other situations, such as selecting appropriate limits for conventional QC [110,111].

4.7. The three-level model

A problem with the general application of these analytical goals is that the available methods may not achieve the required precision and accuracy. For example, those measurands subject to tight homeostatic control (e.g. electrolytes) have a TE that produces minimal performance specifications. To allow for some flexibility, three acceptable precision and bias levels have been adopted [16,81].

With the ratio of B to S_B at 0.25, the percentage of the population outside the upper and lower reference limits moves from a symmetric 2.5% at both limits to 1.3% (at the lower end) and 4.4% (at the upper end), leading to an increase in the population outside the reference interval of 0.8% of the group. This equates to a rise of 0.8/5.0% or 16% misclassified. This precision level was called 'desirable' by Fraser et al. [81].

Using similar calculations, if B/S_B becomes 0.375, then the percentages outside the lower and upper reference limits become 1.0% and 3.3%, leading to a 1.7/5.0 or 34% more misclassified above the reference limit. This level of bias was termed minimum performance.

Lastly, if B/S_B becomes 0.125, then the lower and upper reference limits become 1.8% and 3.3%, which causes a misclassification error of 2%. This was termed optimal performance.

The impact of different precision factors has also been considered. If the ratio of S_A/S_B is 0.5, the Tonks number, then an additional 12% of the variability is added to the TE function. This is termed desirable precision. With a ratio of S_A/S_B of 0.75, there is an increase in variability of 25% (minimal), whereas if S_A/S_B is 0.25, then only 3% (optimal) is added.

Typically a z value for 95% confidence interval is used. However, it can be increased to a 99% confidence interval in some instances, illustrating that all the components of this equation are variables [16].

The Index of Individuality demonstrated that it is best to use intra-individual reference intervals for some measurands when setting performance goals for these same assays. Thus different specifications are suggested for screening, diagnosis, and monitoring. However, in practice, none of the plans may be unachievable by current state-of-the-art procedures.

4.8. Oosterhuis and Sandberg model

Work continues on the development of a TE model (equation (5)) to overcome some of the issues identified by Oosterhuis and Sandberg, including [112]:

- the bias term is applied in monitoring, although this expression had come from the diagnosis-based reference value model.
- Both the maxima of allowable bias and precision are used.

Their approach is based on the same principles but with an accurate calculation of reference limits or reference change values, which leads to more realistic goals for analytes with a high analytical variation relative to the biological variation. Mathematically, this model is an adaptation of Gowan's model and the reference change values' model. With the two underlying principles:

1. The model describes the maximum bias and precision allowable that still maintains the validity of reference values (or reference change values in case of monitoring).
2. The reference limits are defined by both biological and analytical variation.

A key difference is introducing a temporal aspect of the analytical variation, CV_{A0} , from the time ($t = 0$) the reference limits were determined or validated.

In the situation of diagnosis, the CV of the reference value is defined as:

$$CV_{ref} = \sqrt{(CV_G^2 + CV_I^2 + CV_{A0}^2)}$$

The actual (total) variation of test results in a reference population is based on biological variation and CV_A (CV actual analytical):

$$CV_{total} = \sqrt{(CV_G^2 + CV_I^2 + CV_A^2)}$$

Using Gowan's criteria, a maximum of 4.6% of the test results outside a reference limit is considered acceptable.

By including CV_{A0} in the model, it focuses on any increase in CV_A relative to CV_{A0} . The rise in CV_A rather than the absolute value of CV_A determines compliance with quality goals. This model also starts with a precision Gaussian distribution of $CV = \sqrt{(CV_B^2 + CV_{A0}^2)}$ to which is then added any additional analytical variation and bias. This contrasts with Gowan's model, where a Gaussian distribution is assumed with $CV = CV_B$ with reference limits at the point where 2.5% of the results are outside the boundaries. Analytical variation (or analytical variation in combination with bias) is then added to the model with a limit of 4.6% test results outside the reference limits. Hence in this model, there is a 4.4% misclassification.

Thus, the model becomes for diagnosis:

$$TE = 0.275 * CV_{TO} + 1.65 * CV_{A0} \quad (6)$$

where $CV_A = CV_{A0}$

$$CV_{TO} = \sqrt{(CV_G^2 + CV_I^2 + CV_{A0}^2)} \text{ and } Bias_{max} = -1.68 * CV_{TO} + 1.96 * CV_{TO} = 0.275 * CV_{TO}$$

The bias sum terms come from the 2.5% and 4.4% limits of the Gaussian distribution.

4.9. The variance model

The variance model uses squared components because this weighs outliers more heavily than data closer to the mean, and it prevents differences above the mean from canceling out those below.

There are two main models for the combination of random and systematic error based on variances:

- the classical variance model suggested by Harris is $S_{total}^2 = S_{within-lab}^2 + |bias|^2$ [85,113], where s is the within-lab variation, and the laboratory-bias term is considered as a random variable (between-laboratory variance) to combine the two concepts in a variance model. This model has mainly been used in relation to the description of within- and between-individual variation, whether for biological or analytical variation.
- The measurement uncertainty model introduced as an accreditation model (ISO 17025) and described by EURACHEM [114] uses the variance model above to encompass the concepts of GUM (Guide to Uncertainty in measurement) [115] by combining known and assumed uncertainties (type A and type B uncertainties). After the correction of all biases and introducing these corrections' uncertainties, these uncertainties – described as standard deviations or coefficients of variation – are combined as variances.

4.10. Problems with linear TE models

A fundamental problem in the estimation of analytical quality specifications is the combination of random and systematic errors. The nature of random errors is different from systematic errors, the former described by the standard deviation (or coefficients of variation), whereas the latter by differences (from defined target values). 'It would be logical to keep these two incommensurable concepts separated, but there are a considerable desire and pressure for creating such models [2].

Bias is a problem with these models. The concept of the 'true value' has been abandoned in metrology. If an actual value cannot be known, TE cannot be estimated. We must use 'surrogate true values' to estimate TE, but the TE models do not appreciate bias estimation/correction uncertainty. This has been addressed by Frenkel et al. [116].

5. Quality control and external quality assurance

5.1. Sigma metrics

The sigma metric ($\sigma = \{TE - B\}/SD_a$) has been used extensively in Industry and lately in laboratory medicine for comparing methods [117–123]. Capability, $Cp_a = TE/CV_a$ (equivalently TE/SD_a), is an objective measure of an assay's ability to meet pre-defined (user) requirements. Clearly, Cp_a is the sigma metric when no bias is assumed. Note that the sigma metric is derived from the TE, so the sigma metric becomes a TE metric in a sense [124].

In Industry, the sigma metric is data based on the original tolerance limits based on customer expectations, not a TE model. In the application of sigma to laboratory medicine, the concept has further been modified to include a bias term in the numerator, which can be problematic [125–128]. The assumption of zero bias is valid in many routine clinical chemistry tests. It is also in an EQA program if the result is compared to the method or group performance.

The sigma or Capability metric indicates the number of standard deviations of the process inside the analytical goals, so the higher the value, the better. Conventional QC relies on using manufactured control material and statistically-derived control rules to identify assay failure caused by a change in reagents, instrument settings, or calibrator change. The laboratory should select an appropriate QC algorithm to ensure good error detection of critical-sized shifts in performance. Where analytical precision is low relative to performance goals, good error detection is hard to achieve regardless of the QC algorithm used [129]. There are algorithms available for selecting the 'best' QC rules and frequency of QC samples based on using sigma metrics [117,130].

Any QC strategy aims to have well-defined QC rules, which will have at least a probability of error detection (P_{ed}) of 0.9. Westgard and Westgard [118] have constructed sigma quality control assessment tools giving QC rules and QC sample frequencies for assays with Cp_a of 4, 5, and 6 with a P_{ed} of 0.9. Less capable assays require more complex sets of QC rules and a higher frequency of QC samples. The following are examples of some different rules for different capabilities [130].

- $Cp_s < 4$ multi-rule or $1_{2.5s}$ with $n = 4$ QC samples per run (2, twice per shift)
- $Cp_s 4-6$ multi-rule or $1_{2.5s}$ with $n = 2$ QC samples once per shift.
- $Cp_s > 6$ 1_{3s} rule with $n = 2$ QC samples once per shift.

5.2. Patient based quality control (PBRTQC) and quality assurance

We will investigate the change in how QC and EQA are evolving to monitor changes in the patient population parameters directly.

Now we will consider PBRTQC techniques. As stated above, QC in clinical chemistry evolved using proxy materials to identify out of control situations. It is noteworthy that, like sigma metrics to laboratory medicine, this was not the usual situation in Industry. The QC process would be to test the product to ensure it meets customer expectations. As analysers and methods become more reliable, laboratories tended to run fewer QC samples, sometimes as few as one per day (Rosenbaum et al., 2018). This increases the number of patient samples released before it becomes evident that a systematic error has occurred. With the demands on laboratories for high turnaround and large patient volumes, how labs may deal with failure and when they could or would repeat samples during the failure period are concerning [131]. It is now more apparent that there are problems with conventional QC including non-commutability of the sample, non-Gaussian nature of the distribution of QC values, low cost, and insensitivity of the QC rules, particularly for assays with low sigma (≤ 4) [132–136]. The model that we have used for many years in QC is still based on the fact that there are three types of error systematic, random, and proportional, although the last is usually subsumed into the first [72]. However, it is now apparent that this was an oversimplification and that errors can be hybrids of these; they don't follow this model at all [137]. There are other problems with conventional quality control that go to the core of the approach, including its statistical basis [132].

PBRTQC methods have been used since the seventies in hematology analysers. Still, there has been little uptake in clinical chemistry because of a lack of computing capacity or understanding of the laboratory's patient populations being served. As the capabilities of widely used middleware and analyser resident software have advanced, patient-based QC techniques are being adopted. The timing of these programs' availability is fortuitous as the consolidation of laboratories and a large number of screening tests performed on community patients leads to higher testing volumes that presumably represent a more stable population [133].

PBRTQC methods use the population (or subpopulation after removal/exclusion of outliers) reference interval as the control limits [138,139]. Other population-based parameters, such as standard deviation and percentage abnormal results, can be used to monitor assay performance [140]. A form of PBRTQC using an average of delta has been proposed, which uses the population delta check average. But this requires a patient base with multiple retesting to work. Hence it is not suitable for a community setting [141].

Given the relative infrequency of conventional QC events, an attractive feature about using patient results as part of a real-time QC strategy is the possibility of detecting out-of-control error conditions earlier than the next scheduled quality control specimen. While the amount of QC information in a single patient result may be relatively small, the amount of information continually grows with each additional patient sample that is tested. This is the interpretation of 'real-time.' Each result is adding to the information. There are some assays where PBRTQC will never be appropriate such as small volume, volatile population (pediatric or high number of abnormal (some hormone, tumour markers) [138].

The other advantage of PBRTQC is that it is patient-centric. You are seeing changes in the clinical impact of erroneous patient results. Having a 3SD error in IQC does not tell us the clinical effect of the error until we retrospectively retest the samples in question or examine the data in some ways. PBRTQC allows us to see shifts in patient results.

For any QC strategy to be functional, there must be clear procedures for the laboratory technologist to follow in rerunning samples

and detect the point in the analytical run where the error becomes apparent. There is now growing awareness that the future of either conventional QC or a PBRTQC approach will involve more generous use of Artificial Intelligence to monitor the process and manage the analyser response to out of control situations, including troubleshooting, recalibration, and rerunning samples [138,142].

5.3. Patient-based external QA

Patient population-based parameters have also been used in external quality assurance. Thienpont et al. [143–145] developed two online tools, ‘Percentiler’ and ‘Flagger,’ based on moving medians and population percentiles. ‘Percentiler’ monitors daily outpatient means/medians to reflect the stability/variation of performance at the individual laboratory level and its peer group. ‘Flagger’ monitors flagging of results against reference intervals or decision limits used in the individual laboratory, as well as at the peer group level. It is complementary to ‘Percentiler’ in that it directly translates the effect of analytical quality/(in)stability on flagging as surrogate medical decision making.

One of the EQA programs’ key roles is to monitor IVD test stability across laboratories and peers/manufacturers, providing post-market surveillance [146]. Many of the problems of EQA are the same as QC, non-commutability of the sample, the inconsistency of acceptability criteria, and frequency of challenge [147]. Using a patient-based sample approach will solve some of these problems.

6. Conclusion

This Review has described the increasing importance of the concepts of biological variation to clinical chemists. The idea of comparison to ‘reference’ is fundamental in measurement. For the biological measurands, that reference is the relevant patient population, a clinical decision point based on a trial or an individual patient’s previous results. The idea of using biological variation to set quality goals was then realised for setting QC and EQA limits. The current phase of BV integration into practice is using PBRTQC and PBQA to detect a change in assay performance.

The progress from the idea of a ‘normal range’ and ‘average man’ to a biological passport to provide personalised reference intervals demonstrates the greater awareness of individual biological variation. Diagnosis and classification based on clinical decision points will always be an essential purpose of laboratory results, but in the future, monitoring of individuals during the ‘well’ phase of their lives will be the norm.

Throughout the study and application of biological variation and its use in defining acceptability has been a tension between empirical and model-based approaches, from Gaussian to nonparametric statistical approaches, from using very well defined ‘healthy’ to large patient populations, from linear models of bias and precision to uncertainty (variance) models. We see with QC, and EQA schemes move away from a model of elegant statistical process control to using patient data to detect the error, from model-based to big data. These changes occur because we have the volumes of patient data available, the computing capability to perform the necessary calculations, and a growing realisation that the models being used are too crude. It may be that just as the data-driven approach of using ANOVA on large patient data sets can identify intra-individual, intra-group, and analytical variation to be used for estimates of Biological Variation, the same data-based approach could be used for EQA and even QC.

We need data to develop better models, but all models will eventually fail. New data analysed at a deeper level cannot be understood using the existing model, as has always been the way [148].

Biological Variation

Understanding why it is so important?

Funding sources

There was no funding of this work.

Declaration of interest

There were no financial or personal interests that could affect this review.

Permissions

There are no permissions required for the content of this review.

Author agreement

I was the sole author of this review.

Declaration of competing interest

There are no conflicts of Interest.

References

- [1] BIPM, JCGM 200:2008, International vocabulary of metrology - basic and general concepts and associated terms (VIM), Bpm (2008) 90, [https://doi.org/10.1016/0263-2241\(85\)90006-5](https://doi.org/10.1016/0263-2241(85)90006-5).
- [2] P.H. Petersen, Importance of the choice of assumptions and models in the estimation of analytical quality specifications, *Accred Qual. Assur.* 7 (2002) 488–493, <https://doi.org/10.1007/s00769-002-0523-6>.
- [3] P.H. Petersen, C.G. Fraser, Strategies to set global analytical quality specifications in laboratory medicine: 10 years on from the Stockholm consensus conference, *Accred Qual. Assur.* 15 (2010) 323–330, <https://doi.org/10.1007/s00769-009-0630-8>.
- [4] D. Burnett, F. Ceriotti, G. Cooper, C. Parvin, M. Plebani, J. Westgard, Collective opinion paper on findings of the 2009 convocation of experts on quality control, *Clin. Chem. Lab. Med.* 48 (2010) 41–52, <https://doi.org/10.1515/CCLM.2010.001>.
- [5] C. Fraser, P. Hyltoft Petersen, C. Ricos, R. Haeckel, Fraser, et al., *Quality Specifications for Imprecision and Inaccuracy Proposed Quality Specifications for the Imprecision and Inaccuracy of Analytical Systems for Clinical Chemistry*, vol. 1, 1992.
- [6] P. Hyltoft Petersen, P. Rustad, Prerequisites for establishing common reference intervals, *Scand. J. Clin. Lab. Invest.* 64 (2004) 285–292, <https://doi.org/10.1080/00365510410006298>.
- [7] P.L. Rustad, P. Felding, A. Lahti, Proposal for Guidelines to Establish Common Biological Reference Intervals in Large Geographical Areas for Biochemical Quantities Measured Frequently in Serum and Plasma, 2004.
- [8] M. Gallagher, L. Mobley, G. Klee, P. Schryver, *The Impact of Calibration Error in Medical Decision Error*, NIST, 2004.
- [9] R. Jansen, N. Jassam, A. Thomas, C. Perich, P. Fernandez-Calle, A.P. Faria, H. Correia, J.H. Barth, C. Weykamp, C. Cobbaert, M. Thelen, C. Ricos, A category 1 EQA scheme for comparison of laboratory performance and method performance: an international pilot study in the framework of the Calibration 2000 project, *Clin. Chim. Acta* 432 (2014) 90–98, <https://doi.org/10.1016/j.cca.2013.11.003>.
- [10] C.G. Fraser, Test result variation and the quality of evidence-based clinical guidelines, *Clin. Chim. Acta* (2004) 19–24, <https://doi.org/10.1016/j.cccn.2003.12.032>.
- [11] C.G. Fraser, *Biological Variation and Quality for POCT*, 2001.
- [12] P.H. Petersen, Quality specifications based on the analysis of effects of performance on clinical decision-making, *Scand. J. Clin. Lab. Invest.* 59 (1999) 517–521.
- [13] P. Hyltoft Petersen, S. Sandberg, C. Fraser, H. Goldschmidt, A model for setting analytical quality specifications and design of control for measurements on the ordinal scale, *Clin. Chem. Lab. Med.* 38 (2000) 545–551.
- [14] C.G. Fraser, Biological variation: a rapidly evolving aspect of laboratory medicine, *J. Lab. Precis. Med.* 2 (2017), <https://doi.org/10.21037/jlpm.2017.06.09>, 35–35.
- [15] C. Ricos, C. Perich, J. Minchinela, V. Álvarez, M. Simón, C. Biosca, M. Doménech, P. Fernández, C. Jiménez, J. Garcia-Lario, F. Cava, Application of biological variation – a review, *Biochem. Med.* 19 (2009) 250–259.
- [16] C.G. Fraser, *Biological Variation: from Principles to Practice*, AACCPress, Washington, 2001.
- [17] C. Fraser, S. Sandberg, *Tietz Textbook of Chemistry and Molecular Diagnostics - Chapter 7 Biological Variation*, sixth ed., Elsevier, St Louis, 2018.
- [18] E. Cotlove, E.K. Harris, G.Z. Williams, Biological and analytic components of variation in long-term studies of serum constituents in normal subjects III. Physiological and medical implications. <https://academic.oup.com/clinchem/article-abstract/16/12/1028/5679435>, 1970.
- [19] E.K. Harris, P. Kanofsky, G. Shakarji, E. Cotlove, Biological and Analytic Components of variation in long-Term Studies of serum constituents in Normal Subjects II. Estimating biological components of variation. <https://academic.oup.com/clinchem/article-abstract/16/12/1022/5679428>, 1970.
- [20] S. Stahl, The evolution of the normal distribution, *Math. Mag.* 79 (2006) 96–113.
- [21] E. Faerstein, W. Winkelstein, Adolphe quelet: statistician and more, *Epidemiology* 23 (2012) 762.
- [22] G. Eknoyan, Adolphe Quelet (1796-1874) - the average man and indices of obesity, *Nephrol. Dial. Transplant.* 23 (2008) 47–51, <https://doi.org/10.1093/ndt/gfm517>.
- [23] H. Poincare, *Thermodynamique*, Georges Carre, Paris, 1892.
- [24] E.A. Murphy, A scientific viewpoint on normalcy, *Perspect. Biol. Med.* 9 (1966) 333–348, <https://doi.org/10.1353/pbm.1966.0011>.
- [25] M.B. Whyte, P. Kelly, The normal range: it is not normal and it is not a range, *Postgrad. Med.* 94 (2018) 613–616, <https://doi.org/10.1136/postgradmedj-2018-135983>.
- [26] J. Hodnett, Proficiency Testing: we all do it - but what do the results mean? *Lab. Med.* 30 (2018) 316–323.
- [27] R. Grasbeck, Saris NE, establishment and use of normal values, *Establ. Use Norm. Values.* 26 (1969).
- [28] R. Gräsbeck, The evolution of the reference value concept, *Clin. Chem. Lab. Med.* 42 (2004) 692–697. <http://nc.ibk.liu.se/nfkk/NORDKEM>.
- [29] J. Williams Roger, *Biochemical Individuality: the Basis for the Genotrophic Concept*, John Wiley and Sons, New York, 1956.
- [30] C. Perich, J. Minchinela, C. Ricos, P. Fernández-Calle, V. Alvarez, M.V. Doménech, M. Simón, C. Biosca, B. Boned, J.V. García-Lario, F. Cava, P. Fernández-Fernández, C.G. Fraser, Biological variation database: structure and criteria used for generation and update, *Clin. Chem. Lab. Med.* 53 (2015) 299–305, <https://doi.org/10.1515/cclm-2014-0739>.
- [31] A. Coskun, F. Braga, A. Carobene, X. Tejedor Ganduxe, A.K. Aarsand, P. Fernández-Calle, J. Díaz-Garzón Marco, W. Bartlett, N. Jonker, B. Aslan, J. Minchinela, B. Boned, E. Gonzalez-Lao, F. Marques-García, C. Perich, C. Ricos, M. Simón, S. Sandberg, Systematic review and meta-analysis of within-subject and between-subject biological variation estimates of 20 haematological parameters, *Clin. Chem. Lab. Med.* (2019), <https://doi.org/10.1515/cclm-2019-0658>.
- [32] C. Ricos, N. Iglesias, J.V. García-Lario, M. Simón, F. Cava, A. Hernández, C. Perich, J. Minchinela, V. Alvarez, M.V. Doménech, C.V. Jiménez, C. Biosca, R. Tena, Within-subject biological variation in disease: collated data and clinical consequences, *Ann. Clin. Biochem.* 44 (2007) 343–352, <https://doi.org/10.1258/000456307780945633>.
- [33] C. Ricos, V. Alvarez, F. Cava, J. Garcia-Lario, A. Hernandez, C. Jimenez, et al, Integration of data derived from biological variation into the quality management system, *Accred Qual. Assur.* 9 (2004) 128–131.
- [34] A. Carobene, Reliability of biological variation data available in an online database: need for improvement, in: *Clin. Chem. Lab. Med.*, Walter de Gruyter GmbH, 2015, pp. 871–877, <https://doi.org/10.1515/cclm-2014-1133>.
- [35] G.G. Fraser, E.K. Harris, Generation and application of data on biological variation in clinical chemistry, *Crit. Rev. Clin. Lab. Sci.* 27 (1989) 409–437, <https://doi.org/10.3109/10408368909106595>.
- [36] T. Røraas, B. Støve, P.H. Petersen, S. Sandberg, Biological variation: the effect of different distributions on estimated within-person variation and reference change values, *Clin. Chem.* 62 (2016) 725–736, <https://doi.org/10.1373/clinchem.2015.252296>.
- [37] E.K. Harris, Effects of intra- and interindividual variation on the appropriate use of normal ranges, *Clin. Chem.* 12 (1974) 1535–1542. <https://academic.oup.com/clinchem/article-abstract/20/12/1535/5677210>.
- [38] F.R. Vogenberg, C.I. Barash, M. Pursel, *Personalized Medicine - Part 1: Evolution and Development into Theranostics*, vol. 35, P T, 2010.
- [39] M. Plebani, G. Lippi, Personalized (laboratory) medicine: a bridge to the future, *Clin. Chem. Lab. Med.* 51 (2013) 703–706, <https://doi.org/10.1515/cclm-2013-0021>.
- [40] K.A. Sikaris, *Physiology and its importance for reference intervals*, *Clin. Biochem. Rev.* 35 (2014) 3–14.
- [41] M. Puspuram, G. Ertaylan, O. Thas, *Computational Science – ICSS 2020* (2020) 334–347, https://doi.org/10.1007/978-3-030-50423-6_12140.
- [42] A. Coskun, M. Serteser, C. Cavusoglu, M. Kilercik, I. Unsal, A New Model to Determine the Personal Reference Interval of Tests in Laboratory Medicine: A Preliminary Study, *MedRxiv*, 2020, <https://doi.org/10.1101/2020.02.07.20020446>, 2020.02.07.
- [43] D.A. Berry, Bayesian statistics and the efficiency and ethics of clinical trials, *Stat. Sci.* 19 (2004) 175–187, <https://doi.org/10.1214/088342304000000044>.
- [44] P.E. Sottas, N. Robinson, O. Rabin, M. Saugy, The athlete biological passport, *Clin. Chem.* 57 (2011) 969–976, <https://doi.org/10.1373/clinchem.2011.162271>.
- [45] P.E. Sottas, N. Baume, C. Saudan, C. Schweizer, M. Kamber, M. Saugy, Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio, *Biostatistics* 8 (2007) 285–296, <https://doi.org/10.1093/biostatistics/kxl009>.

- [46] N. Robinson, M. Saugy, A. Verne, P.E. Sottas, The athlete biological passport: an effective tool in the fight against doping, *Clin. Chem.* 57 (2011) 830–832, <https://doi.org/10.1373/clinchem.2011.162107>.
- [47] P.E. Sottas, M. Saugy, C. Saudan, Endogenous steroid profiling in the athlete biological passport, *Endocrinol Metab. Clin. N. Am.* 39 (2010) 59–73, <https://doi.org/10.1016/j.ecl.2009.11.003>.
- [48] L.M. Lobigs, P.E. Sottas, P.C. Bourdon, Z. Nikolovski, M. El-Gingo, E. Varamenti, P. Peeling, B. Dawson, Y.O. Schumacher, A step towards removing plasma volume variance from the Athlete's Biological Passport: the use of biomarkers to describe vascular volumes from a simple blood test, *Drug Test. Anal.* 10 (2018) 294–300, <https://doi.org/10.1002/dta.2219>.
- [49] P.E. Sottas, N. Robinson, S. Giraud, F. Taroni, M. Kamber, P. Mangin, M. Saugy, Statistical classification of abnormal blood profiles in athletes, *Int. J. Biostat.* 2 (2006), <https://doi.org/10.2202/1557-4679.1011>.
- [50] P.E. Sottas, C. Saudan, C. Schweizer, N. Baume, P. Mangin, M. Saugy, From population- to subject-based limits of T/E ratio to detect testosterone abuse in elite sports, *Forensic Sci. Int.* 174 (2008) 166–172, <https://doi.org/10.1016/j.forsciint.2007.04.001>.
- [51] M. Saugy, C. Lundby, N. Robinson, Monitoring of biological markers indicative of doping: the athlete biological passport, *Br. J. Sports Med.* 48 (2014) 827–832, <https://doi.org/10.1136/bjsports-2014-093512>.
- [52] A.R. Verne, The athlete biological passport: an integral element of innovative strategies in antidoping, *Br. J. Sports Med.* 48 (2014) 817–819, <https://doi.org/10.1136/bjsports-2014-093560>.
- [53] J. Tyler, S.W. Choi, M. Tewari, Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine, *Curr. Opin. Struct. Biol.* 20 (2020) 17–25, <https://doi.org/10.1016/j.coisb.2020.07.001>.
- [54] K. Guk, G. Han, J. Lim, K. Jeong, T. Kang, E.K. Lim, J. Jung, Evolution of wearable devices with real-time disease monitoring for personalized healthcare, *Nanomaterials* 9 (2019) 1–23, <https://doi.org/10.3390/nano9060813>.
- [55] M. Panteghini, F. Ceriotti, G. Jones, W. Oosterhuis, M. Plebani, S. Sandberg, Strategies to define performance specifications in laboratory medicine: 3 years on from the Milan Strategic Conference, *Clin. Chem. Lab. Med.* 55 (2017) 1849–1856, <https://doi.org/10.1515/cclm-2017-0772>.
- [56] S. Sandberg, C.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, P.H. Petersen, H. Schimmel, K. Sikaris, M. Panteghini, Defining analytical performance specifications: consensus statement from the 1st strategic conference of the European Federation of clinical chemistry and laboratory medicine, in: *Clin. Chem. Lab. Med.*, Walter de Gruyter GmbH, 2015, pp. 833–835, <https://doi.org/10.1515/cclm-2015-0067>.
- [57] S. Lord, L. Irwig, R. Simes, When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann. Intern. Med.* 144 (2006) 850–855.
- [58] A.R. Horvath, S.J. Lord, A. StJohn, S. Sandberg, C.M. Cobbaert, S. Lorenz, P.J. Monaghan, W.D.J. Verhagen-Kamerbeek, C. Ebert, P.M.M. Bossuyt, From biomarkers to medical tests: the changing landscape of test evaluation, *Clin. Chim. Acta* 427 (2014) 49–57, <https://doi.org/10.1016/j.cca.2013.09.018>.
- [59] H. Petersen, L. Jørgensen, I. Brandslund, N. Olivarius, M. Stahl, Consequences of bias and imprecision in measurements of glucose and HbA1c for the diagnosis and prognosis of diabetes mellitus, *Scand. J. Clin. Lab. Invest. Suppl.* 240 (2005) 51–60.
- [60] S. Pasqualetti, F. Braga, M. Panteghini, Pre-analytical and analytical aspects affecting clinical reliability of plasma glucose results, *Clin. Biochem.* 50 (2017) 587–594, <https://doi.org/10.1016/j.clinbiochem.2017.03.009>.
- [61] J.H. Chai, S. Ma, D. Heng, J. Yeong, W.Y. Lim, S.A. Toh, T.P. Loh, Impact of analytical and biological variations on classification of diabetes using fasting plasma glucose, oral glucose tolerance test and HbA1c, *Sci. Rep.* 7 (2017), <https://doi.org/10.1038/s41598-017-14172-8>.
- [62] B. Karon, J. Boyd, G. Klee, Glucose meter performance criteria for tight glycaemic control estimated by simulation modeling, *Clin. Chem.* 56 (2010) 1091–1097.
- [63] M.R. Langlois, M.J. Chapman, C. Cobbaert, S. Mora, A.T. Remaley, E. Ros, G.F. Watts, J. Borén, H. Baum, E. Bruckert, A. Catapano, O.S. Descamps, A. Von Eckardstein, P.R. Kamstrup, G. Kolovou, F. Kronenberg, A. Langsted, K. Pulkki, N. Rifai, G. Sypniewska, O. Wiklund, B.G. Nordestgaard, Quantifying atherogenic lipoproteins: current and future challenges in the era of personalized medicine and very low concentrations of ldl cholesterol. a consensus statement from EAS and EFLM, *Clin. Chem.* 64 (2018) 1006–1033, <https://doi.org/10.1373/clinchem.2018.287037>.
- [64] A. Lyon, P. Kavsak, O. Lyon, A. Worster, M. Lyon, Simulation models of misclassification error for single thresholds of high-sensitivity cardiac troponin I due to assay bias and imprecision, *Clin. Chem.* 63 (2017) 585–592.
- [65] P. Sheehan, J. Blennerhassett, S. Vasikaran, Decision limit for troponin I and assay performance, *Ann. Clin. Biochem.* 39 (2002) 231–236.
- [66] M. Panteghini, Laboratory and Clinical Issues Affecting the Measurement and Reporting of Cardiac Troponins: A Guide for Clinical Laboratories. Chapter 8: Quality Requirements for Troponin Assays - an Overview, Australian Association of Clinical Biochemists, Sydney, 2012.
- [67] L.P. Staub, S.J. Lord, R.J. Simes, S. Dyer, N. Houssami, R.Y.M. Chen, L. Irwig, Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation, *BMC Med. Res. Methodol.* 12 (2012), <https://doi.org/10.1186/1471-2288-12-12>.
- [68] J. Boyd, D. Bruns, Performance requirements for glucose assays in intensive care units, *Clin. Chem.* 60 (2014) 1463–1465.
- [69] D. Bruns, Laboratory related outcomes in healthcare, *Clin. Chem.* 47 (2001) 1547–1552.
- [70] T.D. Christensen, Doctor OF medical science Danish medical bulletin, *Dan. Med. Bull.* 58 (2011). <http://www.sst.dk>.
- [71] A.R. Horvath, P.M.M. Bossuyt, S. Sandberg, A.S. John, P.J. Monaghan, W.D.J. Verhagen-Kamerbeek, L. Lennartz, C.M. Cobbaert, C. Ebert, S.J. Lord, Setting analytical performance specifications based on outcome studies - is it possible?, in: *Clin. Chem. Lab. Med.* Walter de Gruyter GmbH, 2015, pp. 841–848, <https://doi.org/10.1515/cclm-2015-0214>.
- [72] J.O. Westgard, R.N. Carey, S. Wold, Criteria for judging precision and accuracy in method development and evaluation. <https://academic.oup.com/clinchem/article-abstract/20/7/825/5677071>, 1974.
- [73] J. Westgard, Error budgets for quality management: practical tools for planning and assuring the analytical quality of testing processes, *Clin. Lab. Manag. Rev. Offic. Publ.* 10 (1996) 377–384.
- [74] J. Westgard, S. Westgard, Total analytic error; from concept to application, *Clin. Lab. News.* (2013) 1–6.
- [75] G. Klee, Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity, *Clin. Chemistry.* 39 (1993) 1514–1518.
- [76] A.K. Stroobants, H.M.J. Goldschmidt, M. Plebani, Error budget calculations in laboratory medicine: linking the concepts of biological variation and allowable medical errors, in: *Clin. Chim. Acta.* Elsevier, 2003, pp. 169–176, [https://doi.org/10.1016/S0009-8981\(03\)00181-5](https://doi.org/10.1016/S0009-8981(03)00181-5).
- [77] M. Mackay, G. Hegedus, T. Badrick, Assay stability, the missing component of the error budget, *Clin. Biochem.* 50 (2017) 1136–1144, <https://doi.org/10.1016/j.clinbiochem.2017.07.004>.
- [78] J.S. Krouwer, Setting performance goals and evaluating total analytical error for diagnostic assays. <https://academic.oup.com/clinchem/article-abstract/48/6/919/5641700>, 2002.
- [79] C.G. Fraser, P. Hyltoft Petersen, M. Lytken, Setting analytical goals for random analytical situations, *Clin. Chem.* 36 (1990) 1625. <https://academic.oup.com/clinchem/article-abstract/36/9/1625/5648682>.
- [80] C.G. Fraser, H. Petersen, The Importance of Imprecision, 1991.
- [81] C.G. Fraser, H. Petersen, J.-C. Libeer, C. Riccoli, Proposals for Setting Generally Applicable Quality Goals Solely Based on Biology, 1997.
- [82] C. Fraser, P. Petersen, Analytical performance characteristics should be judged against objective quality specifications, *Clin. Chem.* 45 (1999) 321–323. <https://academic.oup.com/clinchem/article/45/3/321/5643020>.
- [83] F. Elevitch, R. Barnett, K. Borovicheny, Analytical goals in clinical chemistry: their relationship to medical care, *Am. J. Clin. Pathol.* 71 (1979) 624–630.
- [84] K. Linnet, A review on the methodology for assessing diagnostic tests, *Clin. Chem.* 34 (1988) 1379–1386.
- [85] E.K. Harris, Statistical principles underlying analytic goal-setting in clinical chemistry, *Am. J. Clin. Pathol.* 72 (1979) 374–382.
- [86] G.R. Jones, K. Sikaris, J. Gill, G. Jones, Commentary "Allowable Limits of Performance" for External Quality Assurance Programs-An Approach to Application of the Stockholm Criteria by the RCPA Quality Assurance Programs, 2012.
- [87] K. Sikaris, Application of the Stockholm Hierarchy to Defining the Quality of Reference Intervals and Clinical Decision Limits, 2012.
- [88] C. Fraser, The 1999 Stockholm Consensus Conference on quality specifications in laboratory medicine, *Clin. Chem. Lab. Med.* 53 (2015) 837–840.
- [89] D.B. Tonks, A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories, *Clin. Chem.* 9 (1963) 217–233.
- [90] Barnett, Medical significance of laboratory values, *Am. J. Clin. Pathol.* 50 (1968) 671–676.

- [91] K.R. Klee, P.G. GG Schryver, Analytical bias specifications based on the analysis of effects on performance of medical guidelines, *Scand. J. Clin. Lab. Invest.* 59 (1999) 509–512.
- [92] E.K. Harris, T. Yasaka, On the calculation of a “reference change” for comparing two consecutive measurements. <https://academic.oup.com/clinchem/article/29/1/25/5667508>, 1983.
- [93] E.W. Randell, S. Yenice, Delta Checks in the clinical laboratory, *Crit. Rev. Clin. Lab. Sci.* 56 (2019) 75–97, <https://doi.org/10.1080/10408363.2018.1540536>.
- [94] D.H. Ko, H. Il Park, J. Hyun, H.S. Kim, M.J. Park, D.H. Shin, Utility of reference change values for delta check limits, *Am. J. Clin. Pathol.* 148 (2017) 323–329, <https://doi.org/10.1093/AJCP/AQX083>.
- [95] C. Naugler, K. Owens, How useful are delta checks in the 21 st century? A stochastic-dynamic model of specimen mix-up and detection, *J. Pathol. Inf.* 3 (2012) 5, <https://doi.org/10.4103/2153-3539.93402>.
- [96] S.H. Park, S.Y. Kim, W. Lee, S. Chun, W.K. Min, New decision criteria for selecting delta check methods based on the ratio of the delta difference to the width of the reference range can be generally applicable for each clinical chemistry test item, *Ann. Lab. Med.* 32 (2012) 345–354, <https://doi.org/10.3343/alm.2012.32.5.345>.
- [97] R.Z. Tan, C. Markus, T.P. Loh, An approach to optimize delta checks in test panels – the effect of the number of rules included, *Ann. Clin. Biochem.* 57 (2020) 215–222, <https://doi.org/10.1177/0004563220904749>.
- [98] D. Stöckl, H. Baadenhuijsen, C.G. Frasei, J.-C. Libeer, C. Ricós, Desirable Routine Analytical Goals for Quantities A Ssayed in Serum Discussion Paper from the Members of the External Quality Assessment (EQA) Working Group A 1) on Analytical Goals in Laboratory Medicine, 1995.
- [99] B.I. Hyltoft Petersen P, de F.O. Jørgensen, L.G.M. Stahl M, B.-J.K. N, Evaluation of systematic and random factors in measurements of fasting plasma glucose as the basis for analytical quality specifications in the diagnosis of diabetes. 3. Impact of the new WHO and ADA recommendations on diagnosis of diabetes mellitus, *Scand. J. Clin. Lab. Invest.* 61 (2001) 191–204.
- [100] G.R. Jones, Measurement uncertainty for clinical laboratories-A revision of the concept, *Clin. Chem. Lab. Med.* 54 (2016) 1303–1307, <https://doi.org/10.1515/cclm-2016-0311>.
- [101] E.M.S. Gowans, P.H. Petersen, O. Blaabjerg, M. Hørder, Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area, *Scand. J. Clin. Lab. Invest.* 48 (1988) 757–764, <https://doi.org/10.3109/00365518809088757>.
- [102] H. Solberg, The concept of reference intervals, *Clin. Chim. Acta* 165 (1987) 111–118.
- [103] D. Stamm, New Concept for Quality Control of Clinical Laboratory Investigations A New Concept for Quality Control of Clinical Laboratory Investigations in the Light of Clinical Requirements and Based on Reference Method Values, 1982.
- [104] C. Fraser, H. Petersen, C. Ricós, R. Haeckel, Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry, *Eur. J. Clin. Chem. Clin. Biochem.* 30 (1992) 311–317.
- [105] P.H. Petersen, M. Hørder, Influence of analytical quality on test results, *Scand. J. Clin. Lab. Invest.* 52 (1992) 65–87, <https://doi.org/10.1080/00365519209104641>.
- [106] C.G. Fraser, P.H. Petersen, Quality goals in external quality assessment are best based on biology, *Scand. J. Clin. Lab. Invest.* 53 (1993) 8–9, <https://doi.org/10.1080/00365519309085446>.
- [107] H. Petersen, D. Stöckl, J.O. Westgard, S. Sandberg, K. Linnet, L. Thienpont, Models for Combining Random and Systematic Errors. Assumptions and Consequences for Different Models, 2001.
- [108] M. Lytken Larsen, C.G. Fraser, P. Hyltoft Petersen, A Comparison of Analytical Goals for Haemoglobin A1, Assays Derived Using Different Strategies, 1991.
- [109] L. Skendzel, R. Barnett, R. Platt, Medical useful criteria for analyte performance of laboratory tests, *Am. J. Clin. Pathol.* 83 (1985) 200–205.
- [110] K. Linnet, 6.1.2.4 analytical goals for accuracy and precision of Plasma creatinine determinations evaluated by Reference method measurements, *Ups. J. Med. Sci.* 98 (1993) 349–354, <https://doi.org/10.3109/03009739309179331>.
- [111] G. Cooper, N. DeJonge, S. Ehrmeyer, J. Yundt-Pacheco, R. Jansen, C. Ricós, M. Plebani, Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality, *Clin. Chem. Lab. Med.* 49 (2011) 793–802, <https://doi.org/10.1515/CCLM.2011.149>.
- [112] W.P. Oosterhuis, S. Sandberg, Proposal for the modification of the conventional model for establishing performance specifications, in: *Clin. Chem. Lab. Med.*, Walter de Gruyter GmbH, 2015, pp. 925–937, <https://doi.org/10.1515/cclm-2014-1146>.
- [113] E. Harris, Proposed goals for analytical precision and accuracy in single-point diagnostic testing. Theoretical basis and comparison with data from the College of American Pathologists proficiency surveys, *Arch. Pathol. Lab Med.* 112 (1988) 416–420.
- [114] U. Magnusson, B and Örnemark, Eurachem Guide: the Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2014.
- [115] JCGM, Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement, Paris, 2008.
- [116] R. Frenkel, I. Farrance, T. Badrick, Bias in analytical chemistry: a review of selected procedures for incorporating uncorrected bias into the expanded uncertainty of analytical measurements and a graphical method for evaluating the concordance of reference and test procedures, *Clin. Chim. Acta* 495 (2019) 129–138, <https://doi.org/10.1016/j.cca.2019.03.1633>.
- [117] J. Westgard, Useful measures and models for analytical quality management in medical laboratories, *Clin. Chem. Lab. Med.* 54 (2016) 223–233, <https://doi.org/10.1515/cclm-2015-0710>.
- [118] J.O. Westgard, S.A. Westgard, Establishing evidence-based statistical quality control practices, *Am. J. Clin. Pathol.* 151 (2019) 364–370, <https://doi.org/10.1093/AJCP/AQY158>.
- [119] R. Hawkins, Managing the pre- and post-analytical phases of the total testing process, *Ann. Lab. Med.* 32 (2012) 5–16, <https://doi.org/10.3343/alm.2012.32.1.5>.
- [120] J.O. Westgard, H. Bayat, S.A. Westgard, Planning risk-based SQC schedules for bracketed operation of continuous production analyzers, *Clin. Chem.* 64 (2018) 289–296, <https://doi.org/10.1373/clinchem.2017.278291>.
- [121] J.O. Westgard, S.A. Westgard, Quality control review: implementing a scientifically based quality control system, *Ann. Clin. Biochem.* 53 (2016) 32–50, <https://doi.org/10.1177/0004563215597248>.
- [122] H. Bayat, Selecting multi-rule quality control procedures based on patient risk, *Clin. Chem. Lab. Med.* 55 (2017) 1702–1708, <https://doi.org/10.1515/cclm-2016-1077>.
- [123] A. Coskun, I. Unsal, M. Serteser, Six sigma as a quality management tool: evaluation of performance in laboratory medicine 247 X six sigma as a quality management tool: evaluation of performance in laboratory medicine, n.d, www.intechopen.com.
- [124] K. Hens, M. Berth, D. Armbruster, S. Westgard, Sigma metrics used to assess analytical quality of clinical chemistry assays: importance of the allowable total error (TEa) target, *Clin. Chem. Lab. Med.* 52 (2014) 973–980, <https://doi.org/10.1515/cclm-2013-1090>.
- [125] W.P. Oosterhuis, A. Coskun, Sigma metrics in laboratory medicine revisited: we are on the right road with the wrong map, *Biochem. Med.* 28 (2018), <https://doi.org/10.11613/BM.2018.020503>.
- [126] A. Coskun, M. Serteser, I. Ünsal, Sigma metric revisited: true known mistakes, *Biochem. Med.* 29 (2019) 142–148, <https://doi.org/10.11613/BM.2019.010902>.
- [127] A. Coskun, I. Unsal, M. Serteser, A. Fraterman, Towards standardization of external quality assessment schemes by using bias values based on biological variation, *Accred Qual. Assur.* 14 (2009) 547–552, <https://doi.org/10.1007/s00769-009-0549-0>.
- [128] W.P. Oosterhuis, M.J. Severens, Performance specifications and six sigma theory: clinical chemistry and industry compared, *Clin. Biochem.* 57 (2018) 12–17, <https://doi.org/10.1016/j.clinbiochem.2018.04.002>.
- [129] H.H. Van Rossum, D. Van Den Broek, Design and implementation of quality control plans that integrate moving average and internal quality control: incorporating the best of both worlds, *Clin. Chem. Lab. Med.* 57 (2019) 1329–1338, <https://doi.org/10.1515/cclm-2019-0027>.
- [130] G. Jones, J. Calleja, D. Cheshier, C. Parvin, J. Yundt-Pacheco, M. Mackay, T. Badrick, Collective Opinion Paper on a 2013 AACB Workshop of Experts Seeking Harmonisation of Approaches to Setting a Laboratory Quality Control Policy, 2015.
- [131] P.J. Howanitz, G.A. Tetrault, S.J. Steindel, Clinical laboratory quality control: a costly process now out of control, *Clin. Chim. Acta* 260 (1997) 163–174, [https://doi.org/10.1016/S0009-8981\(96\)06494-7](https://doi.org/10.1016/S0009-8981(96)06494-7).

- [132] A. Katayev, J.K. Fleming, Past, present, and future of laboratory quality control: patient-based real-time quality control or when getting more quality at less cost is not wishful thinking, *J. Precis. Lab. Med.* (2020) 1–7, <https://doi.org/10.21037/jlpm-2019-qc-03>.
- [133] J. Fleming, A. Katayev, Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: *Clin. Biochem.* 48 (2015) 508–513.
- [134] T. Badrick, A. Bietenbeck, M.A. Cervinski, A. Katayev, H.H. Van Rossum, T.P. Loh, Patient-based real-time quality control: review and recommendations, *Clin. Chem.* 65 (2019), <https://doi.org/10.1373/clinchem.2019.305482>.
- [135] T. Badrick, A. Bietenbeck, A. Katayev, H.H. van Rossum, T.P. Loh, M.A. Cervinski, Implementation of patient-based real-time quality control, *Crit. Rev. Clin. Lab. Sci.* (2020), <https://doi.org/10.1080/10408363.2020.1765731>.
- [136] H.H. Van Rossum, Moving average quality control: principles, practical application and future perspectives, *Clin. Chem. Lab. Med.* 57 (2019) 773–782, <https://doi.org/10.1515/cclm-2018-0795>.
- [137] H. Van Rossum, H. Kemperman, Moving average for continuous quality control: time to move to implementation in daily practice? *Clin. Chem.* 63 (2017) 1040–1041, <https://doi.org/10.1373/clinchem.2016.268060>.
- [138] T. Badrick, A. Bietenbeck, A. Katayev, H.H. Van Rossum, M.A. Cervinski, T.P. Loh, Patient-based real time QC, *Clin. Chem.* 66 (2020) 1140–1145, <https://doi.org/10.1093/clinchem/hvaa149/5880074>.
- [139] H.H. Van Rossum, Moving average quality control: principles, practical application and future perspectives, *Clin. Chem. Lab. Med.* 57 (2019), <https://doi.org/10.1515/cclm-2018-0795>.
- [140] J. Liu, C.H. Tan, T. Badrick, T.P. Loh, Moving standard deviation and moving sum of outliers as quality tools for monitoring analytical precision, *Clin. Biochem.* (2017), <https://doi.org/10.1016/j.clinbiochem.2017.10.009>.
- [141] G. Jones, Average of delta: a new quality control tool for clinical laboratories, *Ann. Clin. Biochem.* 53 (2016) 133–140.
- [142] T.P. Loh, T. Badrick, Using next generation electronic medical records for laboratory quality monitoring, *J. Lab. Precis. Med.* 2 (2017), <https://doi.org/10.21037/jlpm.2017.08.06>, 61–61.
- [143] L.M. Thienpont, D. Stöckl, Percentiler and Flagger-low-cost, on-line monitoring of laboratory and manufacturer data and significant surplus to current external quality assessment, *J. Lab. Med.* 42 (2018) 289–296, <https://doi.org/10.1515/labmed-2018-0318>.
- [144] L. De Grande, K. Goossens, L. Van Uytvanghe, D. Stöckl, L. Thienpont, On-line flagging monitoring - a new quality management tool for the analytical phase, *Clin. Chem. Lab. Med.* 53 (2015) 1197–1204.
- [145] K. Van Houcke, H. Stepman, L. Thienpont, T. Fiers, V. Stove, V. Couck, Long-term stability of laboratory tests and practical implications for quality management, *Clin. Chem. Lab. Med.* (2013), 511227–123.
- [146] F. Braga, M. Panteghini, Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies, *Clin. Chim. Acta* 432 (2014) 55–61, <https://doi.org/10.1016/j.cca.2013.11.022>.
- [147] T. Badrick, A. Stavelin, Harmonising EQA schemes the next frontier: challenging the status quo, *Clin. Chem. Lab. Med.* (2020), <https://doi.org/10.1515/cclm-2020-0343>.
- [148] G.E.P. Box, Some problems of statistics and everyday life, *J. Am. Stat. Assoc.* 74 (1979) 1–4.