Clinica Chimica Acta 532 (2022) 13–20

Contents lists available at ScienceDirect

Clinica Chimica Acta

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

Sigma metrics in laboratory medicine: A call for harmonization

Loris Wauthier [a,*], Laura Di Chiaro [a], Julien Favresse [a,b]

[a] Department of Laboratory Medicine, Clinique St-Luc Bouge, Namur, Belgium
[b] Department of Pharmacy, Namur Research Institute for LIfe Sciences, University of Namur, Namur, Belgium

ARTICLE INFO

Keywords:
Sigma metrics
Quality controls
Analytical performance specification
Total error
Bias
Imprecision

ABSTRACT

Background and aim: Sigma metrics are applied in clinical laboratories to assess the quality of analytical processes. A parameter associated to a Sigma > 6 is considered "world class" whereas a Sigma < 3 is "poor" or "unacceptable". The aim of this retrospective study was to quantify the impact of different approaches for Sigma metrics calculation.

Material and methods: Two IQC levels of 20 different parameters were evaluated for a 12–month period. Sigma metrics were calculated using the formula: (allowable total error (TEa) (%) – bias (%))/(coefficient of variation (CV) (%)). Method precision was calculated monthly or annually. The bias was obtained from peer comparison program (PCP) or external quality assessment program (EQAP), and 9 different TEa sources were included.

Results: There was a substantial monthly variation of Sigma metrics for all combinations, with a median variation of 32% (IQR, 25.6–41.3%). Variation across multiple analyzers and IQC levels were also observed. Furthermore, TEa source had the highest impact on Sigma calculation with proportions of Sigma > 6 ranging from 17.5% to 84.4%. The nature of bias was less decisive.

Conclusion: In absence of a clear consensus, we recommend that laboratories calculate Sigma metrics on a sufficiently long period of time (>6 months) and carefully evaluate the choice of TEa source.

1. Introduction

Nowadays, laboratory medicine faces a constant increase in requests and production of analytical data is consequently steadily growing. As such, the management of laboratory errors is crucial and statistical quality control (QC) procedures are key to face present and future challenges while maintaining analytical quality requirements. Despite the development of valuable tools for improving such aspects, QC frequency varies dramatically between laboratories and many still blindly use an empirical 2 standard deviations (SD) rule for QC rejection [1,2]. To rethink and improve QC procedures in clinical laboratories, modern adaptations of the Six Sigma methodology show great potential [3]. Six Sigma finds its roots in the industrial world of the 1980 s and was conceived in an effort to decrease the manufacture of defective products, improve cost-efficiency and reduce variability in processing [4]. Further developed and spread by quality management specialists, the methodology is today applied in industry, business and the healthcare sector [5]. The first mention of Sigma in the context of clinical laboratories was made by Nevalainen *et al.* in 2000 [6] and was rapidly followed by James Westgard [7]. "Six Sigma" strategies were adapted and applied to the reality of clinical laboratories in addition to other indicators to assess the quality of analytical processes. Sigma metrics are today commonly calculated using a simple equation that includes method bias, imprecision through the coefficient of variation (CV), and analytical performance specifications (APS) for allowable total error (TEa) [8]. Usually, the Six Sigma scale runs from 0 to 6 and provides an estimation of the robustness of a process by defining the number of defects that will be produced per million events [9]. Methods that exhibit high Sigma metrics are analytically robust, meaning that more measurement errors can be tolerated before patient results become erroneous. In comparison, low Sigma metrics assays are those that present potential problematic performance. The ideal goal for analytical methods is a Sigma metric of 6, which means that 3.4 defects will be produced per million events ("world class" tests) whereas a Sigma of 3 equals 66,807 defects/million ("marginal" or"poor" tests) [10]. Far from consisting in an additional statistic produced to be documented in laboratory records and to show auditors, Sigma metrics have direct application in the field of laboratory medicine. In practice, Sigma metrics help laboratories to optimize the number of controls, run lengths and the number of QC rules to apply to those controls, enabling them to decrease QC costs,

* Corresponding author.
E-mail addresses: loris.wauthier@icloud.com (L. Wauthier), julien.favresse@slbo.be (J. Favresse).

https://doi.org/10.1016/j.cca.2022.05.012
Received 7 April 2022; Received in revised form 27 April 2022; Accepted 13 May 2022
Available online 17 May 2022

rationalize QC procedures and optimize staff efficiency [3,11].

The main obstacle to a wider use of Sigma metrics in laboratories comes from the lack of consensus regarding the most appropriate data used to determine Sigma. Each variable in the Sigma equation may be extracted from many sources and depends on different factors: the extent of time to consider for CV and bias determination, the material to use for bias determination (i.e., peer comparison program (PCP) that makes use of the routine QC data or external quality assessment program (EQAP)/ proficiency testing program (PTP)) [12] and the APS of the method (i.e., source of TEa) (e.g., CLIA, RiliBAK, EFLM biological variation database). Appropriate guidelines and harmonization are missing and shall therefore help laboratory specialists in implementing Sigma metrics.

The aim of this study was to evaluate the impact of different combinations of data sources for Sigma metrics calculation.

2. Material and methods

2.1. Study design

This retrospective study was conducted for 12 months between January 2020 and December 2020 at the Clinique Saint-Luc Bouge (Namur, Belgium). The Sigma metrics was calculated for 20 parameters based on TEa, bias and imprecision. For that purpose, a total of 67,483 data from 39 different lots of IQC were analyzed.

2.2. Analytical methods

Analyses were performed on a cobas 8000 (Roche Diagnostics, Basel, Switzerland), according to the manufacturer's recommendations. A total of 20 parameters were studied, namely: sodium (Na), potassium (K) (ion-selective electrode (ISE) module); albumin (ALB) (502 module); total cholesterol (TC), HDL–cholesterol (HDL-C), triglycerides (TG), C-reactive protein (CRP), total proteins (TP), creatinine (CREAT), urea (702 module); N-terminal pro-brain natriuretic peptide (NT-proBNP), high-sensitive cardiac troponin T (hs-cTnT), thyroid-stimulating hormone (TSH), free thyroxine (FT4), cortisol (CORT), luteinizing hormone (LH), follicle-stimulating hormone (FSH), estradiol (E2), progesterone (PROG) and prolactin (PRL) (801 module). ALB, CORT, NT-proBNP, LH, FSH, E2, PRL and PROG were analyzed on a single analyzer. CREAT, urea, hs-cTnT, Na, K, TG, TC, HDL, CRP and TP were analyzed on two distinct analyzers. TSH and FT4 were analyzed on three distinct analyzers.

2.3. Sigma metrics calculation

Internal quality controls (IQC) from Bio-Rad (California, USA) were used for all parameters, except hs-cTnT for which IQC were gathered from Roche Diagnostics. Two levels of IQC were evaluated for all parameters (level 1 and 2). Coefficients of variation were calculated for each parameter according to the equation:

$$CV(\%) = \frac{Standard\ deviation}{Mean} x100$$

Raw data for IQC were extracted from the cobas IT Middleware and CV were calculated per month for each analyzer. Outliers were excluded from the analysis because of the effect they may have on the mean (i.e., inversion of levels or aberrant results) [13 14]. The bias was calculated in two different ways. Firstly, the bias was calculated by comparing the mean IQC results of each parameter to the monthly target bias obtained using the QCNet software from Bio-Rad or the eLabQC program from Roche Diagnostics (for hs-cTnT only) (bias 1, PCP). Secondly, the bias was calculated based on the EQC program organized by the Belgian Institute of health (Sciensano). Since EQA samples are not sent every month, yearly data were used for bias calculation (bias 2, EQAP or PTP). The bias was calculated using the following equation:

$$Bias(\%) = \frac{(Mean\ of\ IQC - PCP\ or\ EQAP/PTP)}{PCP\ or\ EQAP/PTP}$$

Nine different APS for TEa were used: (I) Royal College of Pathologists of Australasia (RCPA), (II) Clinical Laboratory Improvement Amendments (2019) (CLIA), Guidelines of the German Federal Medical Council (III) column 3 (i.e., permissible relative deviation of a single result or the relative root mean square, respectively) (RiliBAK3) and (IV) column 5 (i.e., permissible relative deviation in EQA) (RiliBAK5), (V) Randox International Quality Assessment Scheme (RIQAS), (VI) Sciensano, and from the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) biological variation database using (VII) minimal, (VIII) desirable and (IX) optimal specifications. Since the EFLM does not provide biological variation data for hs-cTnT and NT-proBNP, data from the literature were obtained [15,16]. The 18 combinations of data obtained for the CV, bias and TEa permitted the calculation of Sigma metrics for each parameter. Sigma metrics were calculated using the following equation:

$$Sigma\ metric = \frac{TEa(\%) - Bias(\%)}{CV(\%)}$$

categorized as Sigma > 6 ("world class" tests), Sigma 3–6 and Sigma > 6 ("poor" and "unacceptable" tests). A total of 18 ways to calculate the Sigma metrics were presented and compared for 20 parameters and two levels of IQC.

2.4. Statistical analysis

Median and inter-quartile ranges (IQR) were used to present the data. Normality of distribution was tested using the D'Agostino-Pearson test w/wo log transformation. Differences between more than two groups was assessed using an ordinary two-way ANOVA with multiple comparison tests. Data analysis was performed using GraphPad Prism (version 9.3.0) and Microsoft Excel (version 16.58). $p < 0.05$ was used as a significance level.

3. Results

Sigma metrics varied greatly over time and with various bias and TEa combinations (Figs. 1, 2 and **Supplemental Fig. 1. A to R**). Representative examples of the evolution of Sigma metrics for TC and TSH are presented in Fig. 1 and Fig. 2. For TC, important variability from low (<3) to high (>6) Sigma metrics were observed over time for both IQC levels and analyzers (median CV ranging from 31.3% to 48.7%). On the analyzer A, while most combinations (16/18; 88.9%) presented a Sigma metrics > 6 at month 4, Sigma metrics at months 9 and 10 were however all < 6. Considering the analyzer B, the overall variability was less pronounced compared to the analyzer A (48.7% *vs* 31.9% for level 1 and 41.8% *vs* 31.3% for level 2) but Sigma metrics were also highly dependent on the month of calculation. Most Sigma metrics were < 3 for months 3 to 5 while half combinations presented a Sigma metrics > 6 for months 6 to 12. On both analyzers, the TEa source strongly impacted the Sigma metrics. The highest Sigma metrics were observed using TEa from RIQAS, CLIA and RiliBAK5 and the lowest Sigma metrics using TEa from OBV, RCPA and Sciensano. For TSH, high Sigma metrics variations were also observed over time, with the analyzer B presenting a lower variation compared to analyzers A and C. As observed for TC, inter-month variation of Sigma results were important and the nature of bias impacted less the calculation. Sigma metrics were the highest using TEa from MBV and DBV while the lowest ones using TEa from Sciensano and RiliBAK 3 (Fig. 2). For the two representative examples (TC and TSH), the nature of bias did not highly impact Sigma metrics (Figs. 1 and 2).

Considering all other parameters, the variability was higher using level 1 compared to level 2 for ALB, CRP, Na (analyzer A), NT-proBNP, PRL, and TP (analyzer B) (i.e., no overlap of IQR). On the contrary, the

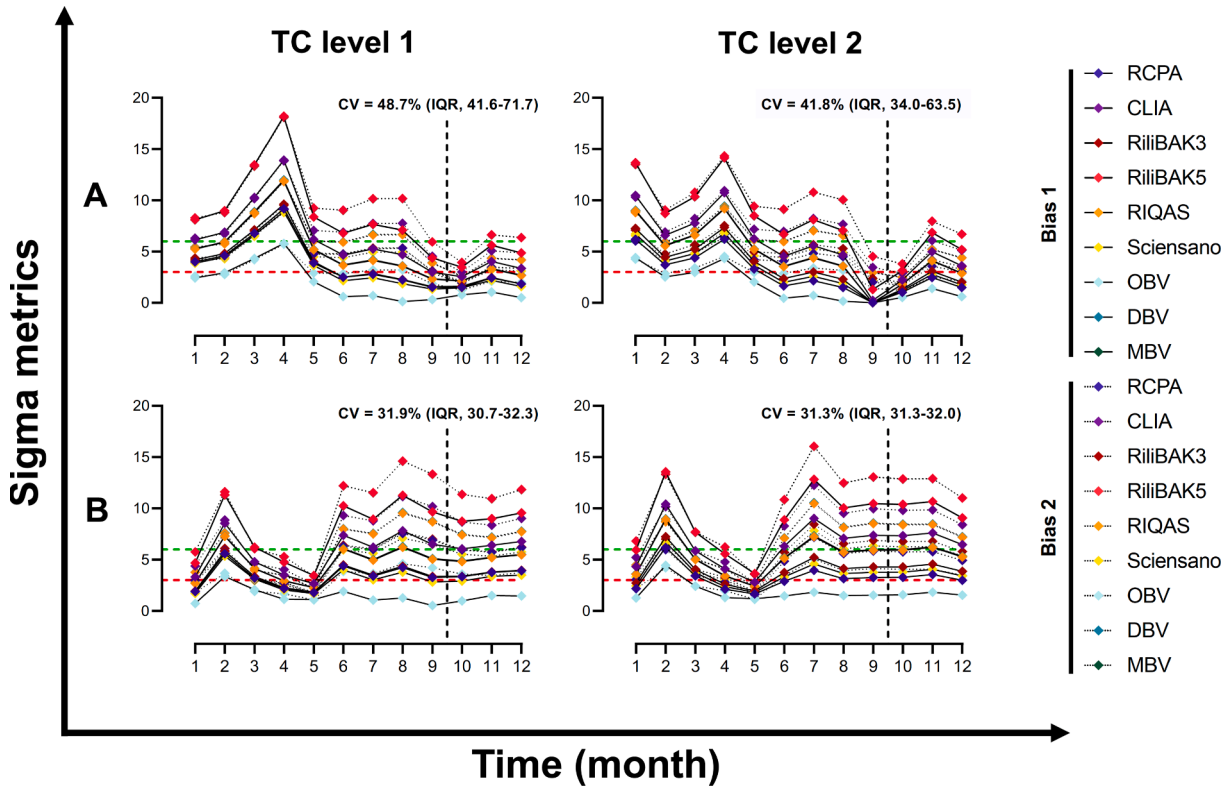

Fig. 1. Evolution of Sigma metrics on a 12-month period using 18 combinations of bias and TEa for TC on 2 analyzers. Plain lines represent combinations using bias 1 and dashed lines represent bias 2. Colors represent TEa sources. Horizontal red dotted line corresponds to 3 Sigma and green dotted line to 6 Sigma. Vertical black dotted line represents a QC lot change. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

variability was higher using level 2 for FSH, hs-cTnT, K, LH, Na (analyzer B), and TP (analyzer A) (**Supplemental Fig. 2. A, D, F, I, J, K, L, M, N, Q**). Regarding the parameters analyzed on two different analyzers, the variability was higher on the analyzer A for CRP (level 2), K (both levels), Na (both levels), and TP (level 2). However, the variability was lower on the analyzer A for CRP (level 1), FT4 (level 2), hs-cTnT (level 1), and TP (level 1) (**Supplemental Fig. 2. D, G, I, J, L, Q**). Overall, we observed a median Sigma metrics variation of 32% (IQR, 25.6–41.3%). CRP (level 1, analyzer B) showed the highest variation over time (78% (IQR, 52.9–83.8)) and CREAT (level 2, analyzer B) the lowest (10.6% (IQR, 10.5–29.4). Our results therefore suggest high variation of Sigma metrics when calculated on a monthly basis.

To overcome this monthly variation, we also decided to calculate annual Sigma metrics (median of monthly Sigma metrics) for all 20 parameters according to bias and TEa combinations (Fig. 3). The heatmap presents differences in Sigma metrics based on: (1) parameters (i.e., rows), (2) bias/TEa combinations (i.e., columns) and (3) IQC levels (Fig. 3). The percentage of Sigma metrics < 3 (%Sigma < 3, blue tone), between 3 and 6 (%Sigma 3–6, beige tone) and > 6 (%Sigma > 6, red tone) was calculated per row and column.

We identified that CREAT and TP mostly have Sigma metrics < 3 (50% to 77.8%) for both IQC levels. ALB, CRP, K and PROG showed>50% of Sigma < 3 with level 1 and NT-proBNP with level 2. On the contrary, CORT, HDL, LH, PRL, TG, hs–cTnT and TSH showed>50% of Sigma > 6 for both levels (50% to 100%). Among them, CORT, LH, PRL and TG all had Sigma metrics > 3.

The proportion of Sigma metrics < 3 and > 6 were 30.8% and 37.3% for level 1 and 18.9% and 50.6% for level 2, respectively (Fig. 3). This shows that Sigma metrics tends to be lower using the level with the lowest concentration.

TEa sources that led to the highest proportion of Sigma metrics > 6 were DBV (50% and 55% for bias 1 and 2), MBV (62.5% and 65.0%), CLIA (60% and 67.5%) and RiliBAK5 (81.3% and 84.4%). Contrariwise, higher proportions of Sigma metrics < 6 were observed using OBV (17.5% and 17.5% for bias 1 and 2), RiliBAK3 (25.0% and 21.9%), Sciensano (25.0% and 27.8%), and RCPA (26.3% and 36.8%). These results are consistent with the TEa repartition per parameter presented in **Supplemental Fig. 1**, where DBV, CLIA, RiliBAK5 and MBV had the highest median TEa whereas OBV, Sciensano and RCPA the lowest median TEa (**Supplemental Fig. 1 and** Fig. 3). Furthermore, we identified that the impact of the bias selection was only moderate. The proportions of Sigma metrics < 3 and > 6 were 28.7% and 41.7% for bias 1 and 21.0% and 46.2% for bias 2, respectively (Fig. 3).

Complementarily, we calculated a median Sigma per parameter using all bias and TEa combinations calculated on the 12-month period for both levels (Fig. 4). Considering level 1, 5 parameters had a median Sigma metric < 3 (25%), 8 between 3 and 6 (40%), and 7 above 6 (35%). Regarding level 2, 2 parameters had a Sigma metric < 3 (10%), 8 between 3 and 6 (40%), and 10 were > 6 (50%). CREAT and TP both had a median Sigma < 3 for levels 1 and 2. Sigma metrics < 3 were also observed for ALB, CRP and PROG for the level 1, and for NT-proBNP considering the level 2. High Sigma metrics (i.e. > 6) were observed for CORT, HDL-C, LH, PRL, TG, and TSH for both IQC levels. These figures further illustrate the major differences in Sigma metrics calculation using different bias and TEa combinations as well as the differences between IQC levels.

4. Discussion

The Six Sigma scale is presented as a powerful and accessible tool to estimate the quality of an analytical process, that includes APS (i.e., TEa) and systematic (i.e., bias) and random (i.e., CV) error estimations [3]. However, there is a multitude of possibilities for data used for its calculation and no clear consensus exists so far. This may be responsible for the lack of application of the Sigma metrics in clinical laboratories [1,2,17]. In the literature, we found several teams that have calculated

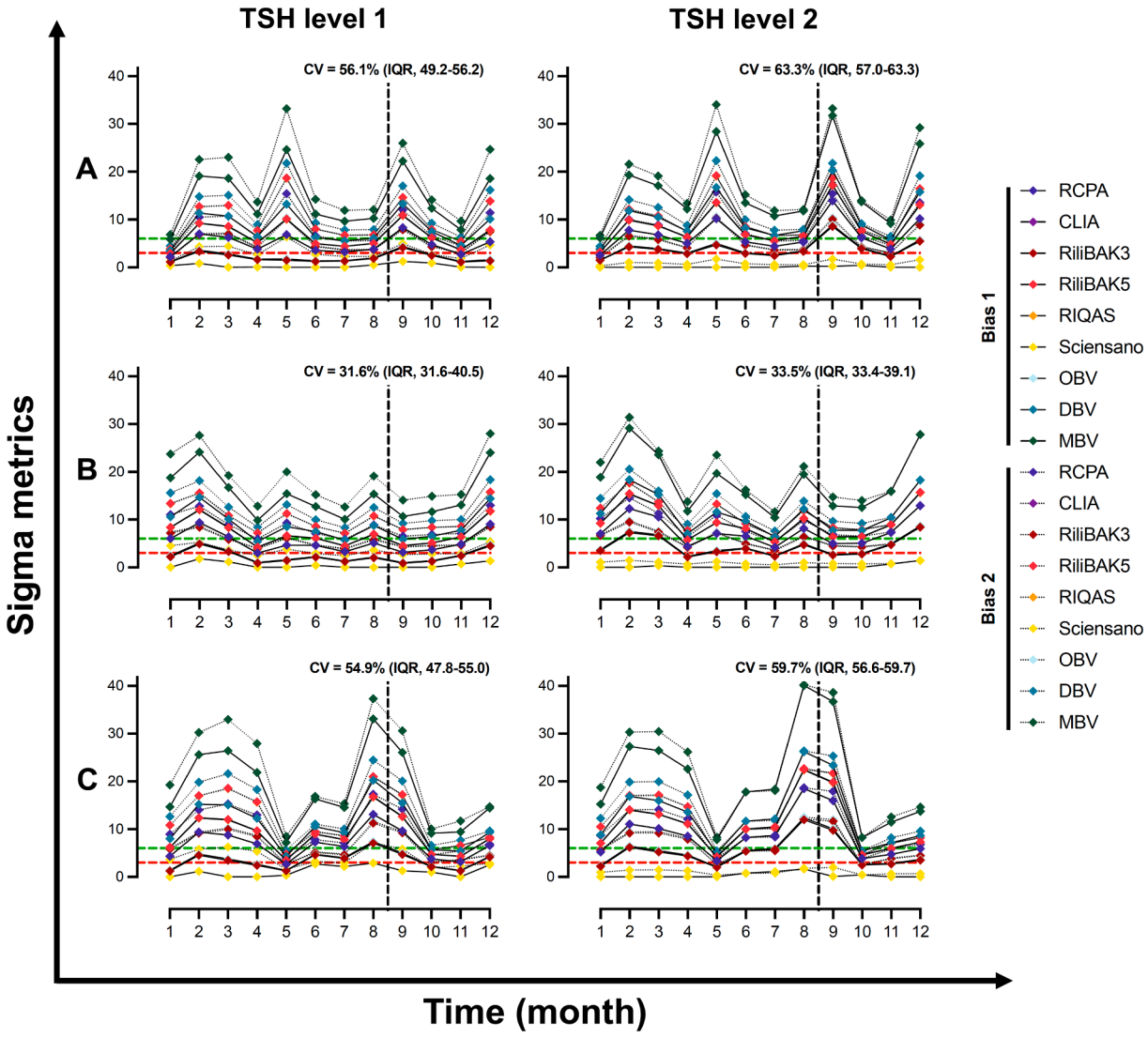

Fig. 2. Evolution of Sigma metrics on a 12-month period using 18 combinations of bias and TEa for TSH on 3 analyzers. Plain lines represent combinations using bias 1 and dashed lines represent bias 2. Colors represent TEa sources. Horizontal red dotted line corresponds to 3 Sigma and green dotted line to 6 Sigma. Vertical black dotted line represents a QC lot change. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sigma metrics [14,17–41]. Many differences across studies were noticeable (nature of bias and of TEa as well as the study length that impact the precision). This reflects well the lack of harmonization that clinical laboratories are facing to apply Sigma metrics. As the choice of calculation combinations may be cause of great variations in Sigma metrics and may lead to substantially different conclusions regarding the quality of a method [13], we decided to address that problematic in the present study.

We first showed that the Sigma calculation was highly depending on timing, as observed in other studies [22,24]. A high inter-month variation was observed for all the 20 parameters evaluated in our study (Figs. 1 and 2, **Supplemental Fig. 2**). Parameters classified as "world class" (Sigma > 6) on a particular month may therefore be classified as "poor" or "unacceptable" (Sigma < 3) in the following months before reaching again a higher Sigma grade (Figs. 1 and 2). Since TEa goals are invariable, variations that occurred over time were only related to changes in bias and CV. This therefore illustrated the instability of performance over time, suggesting that Sigma should be calculated on a larger period (i.e., >6 months) to ensure a more robust estimation of precision and bias, as performed in several studies [23,26,27,29,38,39]). Although we cannot exclude that this phenomenon was due to variability in our own laboratory, this is mostly unlikely. To consider all contributions to long-term variation, CLSI C24 states that imprecision should be estimated from cumulative mean and SD based on several months [42]. A multi-site study might prove useful to confirm our observations.

We also showed that various analyzers for the same parameter might present different Sigma metrics over time. Some proposed to calculate a mean Sigma when the same parameter is analyzed on different analyzers [43]. Furthermore, the CV may also vary across reagent lots, as recently demonstrated [44]. Sigma metrics should be reevaluated in a periodic manner to consider such variability but month-to-month reevaluation will not be practical in clinical laboratories since profound modifications of QC procedures might happen. Additionally, the inclusion of precision data from a short period of time (i.e. <1 month) as performed in some studies [18,25,28,34,37] might falsely present high Sigma metrics. The Sigma metrics also vary depending on the level of IQC considered [28]. Lower Sigma were observed for IQC level 1 compared to level 2 (Fig. 3). This is due to the lower concentration of the parameter in level 1 and then to a higher CV. Whereas some propose to consider CV of different IQC levels to calculate Sigma metrics to apply the strictest rules in case of discrepancies between levels [33,41,43], other proposed to calculate a mean CV between IQC levels to compute an unique Sigma value per parameter [14]. However, and since Sigma results may show values

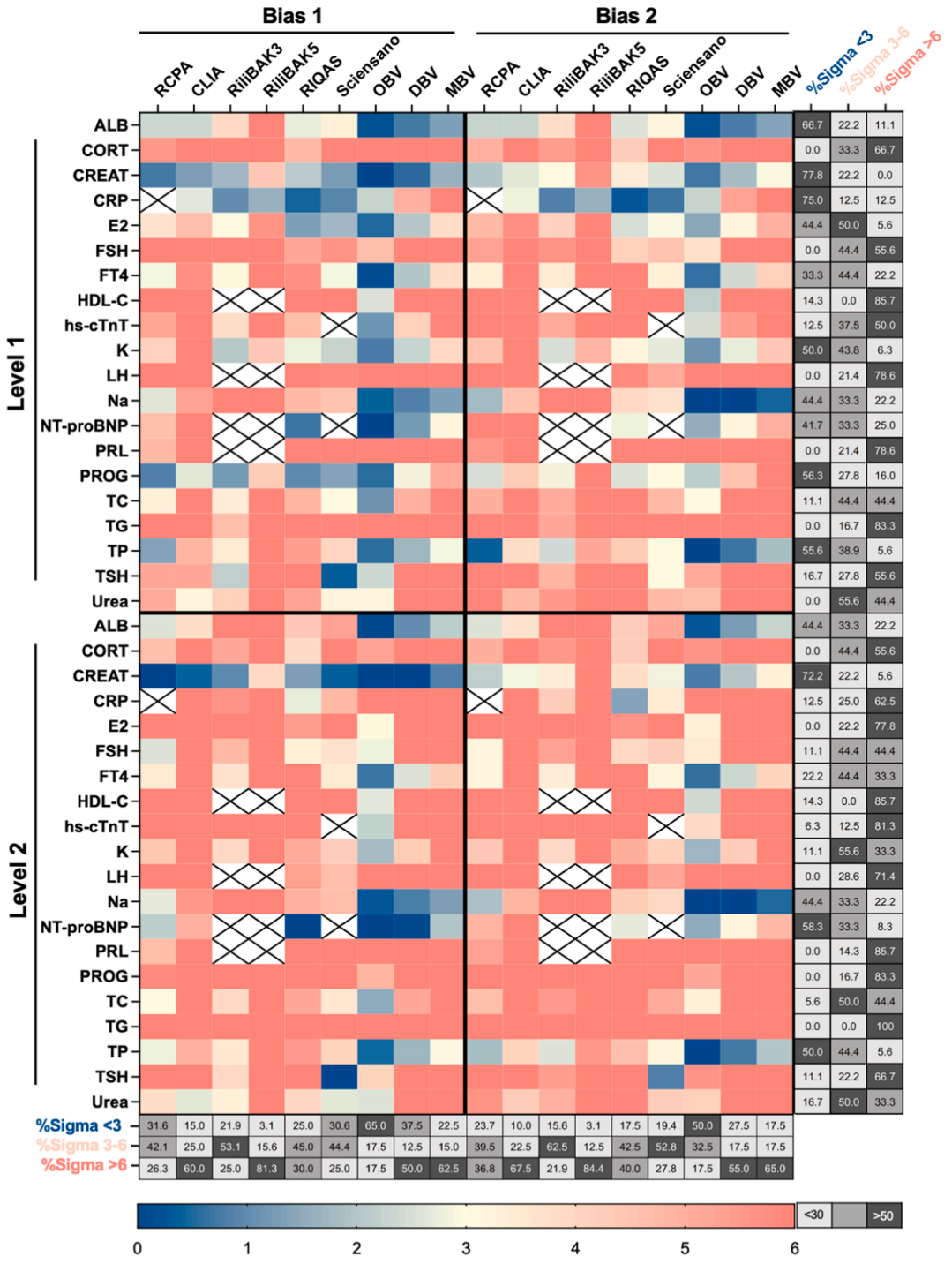

Fig. 3. Heatmap of Sigma metrics for 20 parameters (2 IQC levels) using 18 different bias/TEa sources. The left part of the heatmap corresponds to bias 1 and the right to bias 2. The upper part corresponds to IQC level 1 and the lower part to IQC level 2. Colors represent the Sigma scale from 0 Sigma to 6 Sigma and above. Grey shades correspond to the percentage of combinations in each of 3 Sigma ranges (%Sigma < 3, %Sigma 3–6 and %Sigma > 6). Parameters with unavailable TEa data are represented with a black cross.

above 6, high discrepancies between levels (e.g., level 1 showing Sigma 3 and level 2 Sigma 9), might conclude at a mean high Sigma, ignoring the conclusion that would have been made on the lowest level. It was therefore proposed that Sigma above 6 should be rounded to 6 before calculating the mean, meaning that no average will ever be at 6 Sigma [11]. The expertise of laboratory medicine specialists may also be valuable for the choice of the relevant clinical decision level that should be considered for Sigma metrics calculation.

In our study, we compared two different approaches based on PCP and EQAP (or PTP). These were both available in our laboratory and each of them may be flawed by different characteristics (e.g., low frequency of EQA leading to high uncertainty). As previously suggested [13,43], it appeared that the source of bias impacted less the Sigma metrics (Fig. 3). Some studies reported the use of bias based on the difference with the manufacturer's IQC targets [23,25,32] although this is generally discouraged [13,24]. Others also considered that the bias could be rounded to zero [13].

The last variable of the Sigma equation (i.e. TEa), is a well-known

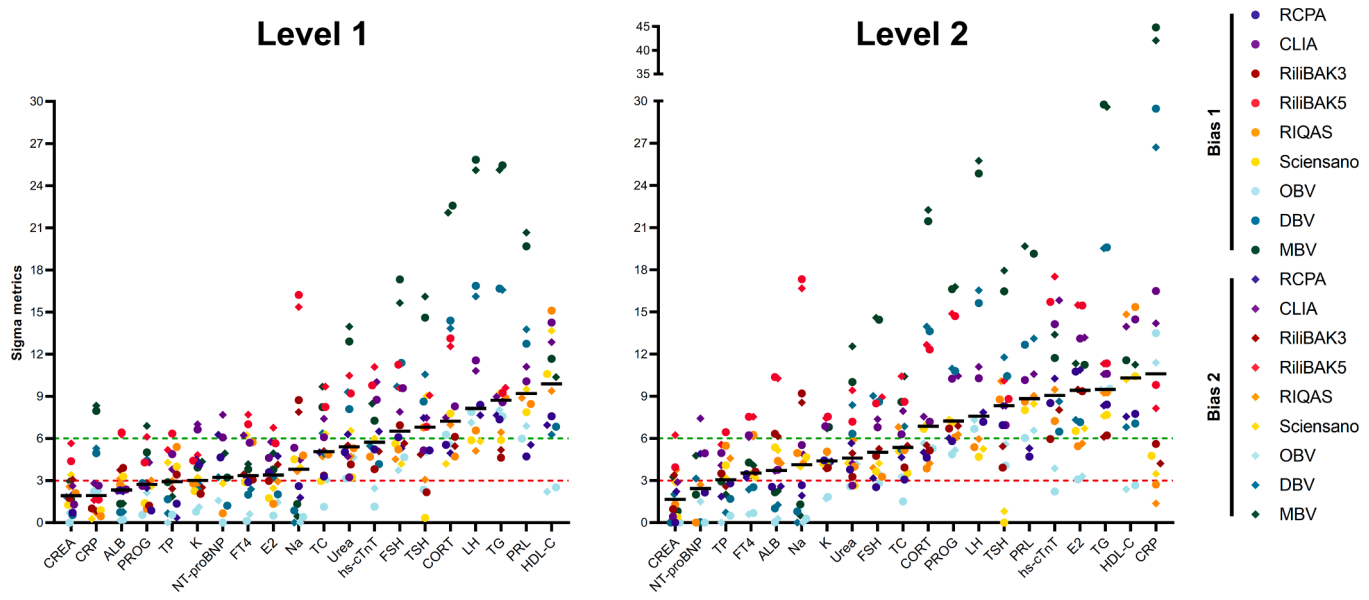

Fig. 4. Median Sigma metrics calculated on all bias/TEa combinations for each parameter ranked by increasing median order. Each symbol corresponds to median Sigma calculated on a 12-month period using a single bias/TEa combination. Red dotted line corresponds to 3 Sigma and green dotted line to 6 Sigma. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

source of variation in Sigma calculation [13,25,28,43,45]. The 2014 Milan strategic conference has proposed models to objectively derive APS based on (1) clinical outcomes, (2) biological variation and (3) state-of-the-art (SOTA), the latter being defined as the highest level of analytical performance technically achievable (i.e. CLIA, RiliBAK, RIQAS, RCPA) [46,47]. Compared to APS based on clinical outcomes that are scarce in the literature [47], biological variation data that cover a wide range of measurands are directly available and updated online on the EFLM biological variation database [48]. References included in meta calculations only are studies that comply with the critical appraisal checklist that has been developed by the EFLM biological variation group [49,50]. The EFLM biological variation database replaces the canonical Ricos database [48]. The variety of TEa sources available for a single parameter represents a major impediment to the application of Sigma metrics [45] and make the performance comparison between laboratories difficult [17]. Our results demonstrated a dramatic variation in Sigma metrics, mostly due to the choice of TEa (Fig. 3). The higher Sigma metrics observed where the ones with the higher TEa. For instance, MBV presented a majority of Sigma metrics > 6, whereas the more restrictive OBV led to lower Sigma metrics (Fig. 3 **and Supplemental Fig. 1**). Other sources such as RCPA, RiliBAK3 and RIQAS performed quite in a similar way in terms of Sigma metrics (Fig. 3 **and Supplemental Fig. 1**). APS vary for many reasons including the strictness of the regulatory regime or the method of determination (i.e., biologically-based or empirical consensus).

Out of 25 studies reviewed, 17 (68.0%) only used 1 TEa source that mostly was based on CLIA 1988 recommendations (10/18; 55.6%) [14,17–19,21–41]. None used the latest CLIA 2019 that are more restrictive compared to 1988, though not yet ratified into law. Few other studies evaluated the impact of TEa on Sigma metrics. Hens *et al.* showed that Sigma metrics tend to be higher using RiliBAK, followed by CLIA and desirable biological data from Ricos [25]. Three other studies also observed higher Sigma metrics using CLIA compared to Ricos or EFLM [22,39,43]. Additionally, higher Sigma metrics using MBV from Ricos followed by CLIA, and OBV from Ricos were observed [39]. These observations are consistent with our observations (Fig. 3, **Supplemental Fig. 1**). Since APS are neither standardized nor harmonized, we must grudgingly accept that benchmarking studies will be similarly afflicted.

An algorithm that considers the hierarchy of the 2014 Milan consensus was proposed by Varela and Pacheco to guide laboratory specialists to evaluate which TEa source better fits the test analytical performance [51]. The authors recommend to use DBV in case of Sigma metrics > 5.15 and to use MBV if the Sigma is < 5.15. The OBV is used in case of Sigma > 12. For that purpose, the bias expressed as a percentage of TEa should also be $< 50\%$. They only propose to use a TEa source of lesser hierarchy in case of low Sigma (i.e., <5.15) when calculated with MBV. They propose to choose a TEa than can bring you to obtain a Sigma metrics > 5.15. In our study, we found that the proportion of Sigma metrics > 6 was the highest for MBV and RiliBAK5. In fact, Sigma metrics were globally lower considering SOTA compared to MBV. In other terms, the strategy proposing to shift from MBV to SOTA in case of low Sigma metrics could not be very useful in clinical practice. The other concern about this algorithm is that it brings users to "force" the use of Sigma metrics > 5.15. Depending on the choice made by the laboratory specialist, the same analytical process may be deemed "poorly performant" while being classified as "excellent" using a second one. It is however a valuable tool to navigate through all existing TEa sources. In our study, the variation across Sigma calculation were so important that it is nearly always possible to find a combination where the Sigma could be > 6 for each of the 20 parameters studied (Fig. 3). The evaluation of the impact of different TEa is therefore important, notably since some OBV and DBV TEa criteria are known to be unachievable with current laboratory platforms [13,25,51]. It should also be noted that the Milan hierarchy of 2014 explicitly acknowledges that no model of APS will apply to all parameters and encourages diversity of choices [46]. Given the constant evolution of testing and clinical treatments (i.e., technical advancements, new application of existing tests, clinical criteria), APS will be dynamic and ever-changing, making it difficult to imagine that any strategic conference may point to a unique set of specifications. The debate over the right APS is not one that can be settled eternally, but instead, only momentarily. Some analytes have clear clinically-based APS (e.g., HbA1c), while others have more appropriate biological variation APS, and those that lack the former types, must settle for the SOTA specifications [47]. Of note, databases do not provide TEa for all parameters, which prevents the use of a single source in a laboratory.

The Sigma metrics will determine the IQC frequency (i.e., number of patients samples between two QC events) and the Westgard rules to apply. This aspect is more harmonized and a new tool is now available online [52,53]. Establishing a form of consensus on the use of APS goals would be a challenging initiative. The issue of Sigma metrics

harmonization in clinical laboratories is nonetheless wider, as the equation for Sigma metrics calculation is the object of a currently unresolved dispute between experts in the field [54,55]. The equation used in this study and generally used in the literature regarding Sigma applications [14,17–41] was indeed adapted from the industry to laboratories by including bias as a variable, which was claimed to be mathematically incorrect and incoherent when compared to the application of Sigma in the industry from which it originated [56–58]. It seems however likely that given the lower impact of bias on Sigma metrics, calculation of Sigma excluding this variable would have produced similar results to those observed here. This further supports that although it may be classified as optimistic or naive because of obvious impediments, harmonization of the approach would benefit laboratory specialists who wish to calculate Sigma metrics and facilitate comparison between laboratories. Even if making the right choices for the calculation of Sigma is complex, we believe that laboratory specialists should not fear its use as this would deprive them of an inestimable ally on their journey to better manage IQC.

In conclusion, Sigma metrics are valuable tools to evaluate the quality of analytical processes and to optimize QC rules and frequency in clinical laboratories. However, the vast choice of sources for precision, bias and APS represents a major impediment to Sigma deployment and harmonization is necessary to ensure adequate use of the metrics. In absence of general guidelines, we recommend that laboratories calculate Sigma metrics on a sufficiently long period of time (>6 months) and carefully evaluate the choice of TEa source. The nature of bias was less decisive for Sigma metrics.

CRediT authorship contribution statement

Loris Wauthier: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Laura Di Chiaro:** Formal analysis, Investigation. **Julien Favresse:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Sten Westgard for his precious comments on the manuscript and insightful discussion on Sigma metrics.

References

[1] Rosenbaum MW, Flood JG, Melanson SEF, Baumann NA, Marzinke MA, Rai AJ, et al. Quality Control Practices for Chemistry and Immunochemistry in a Cohort of 21 Large Academic Medical Centers. Am J Clin Pathol. 2018;150(2):96-104.
[2] WestgardQC. The 2021 European (and others) QC Survey Results [December 19, 2021]. Available from: https://www.westgard.com/2021-europe-et-al-qc-survey-results.htm.
[3] S. Westgard, H. Bayat, J.O. Westgard, Analytical Sigma metrics: A review of Six Sigma implementation tools for medical laboratories, Biochem Med (Zagreb). 28 (2) (2018), 020502.
[4] Harry M SR. Six Sigma: The breakthrough management strategy revolutionizing the world's top corporations. 1st ed. ed. New York 2000.
[5] S. Bisgaard, Quality Management and Juran's Legacy, Quality Engineering. 20 (4) (2008) 390–401.
[6] D. Nevalainen, L. Berte, C. Kraft, E. Leigh, L. Picaso, T. Morgan, Evaluating laboratory performance on quality indicators with the six sigma scale, Arch Pathol Lab Med. 124 (4) (2000) 516–519.
[7] J.O. Westgard, Six Sigma QC Design and Control, Madison QC Westgard (2001).
[8] J.M. Gras, M. Philippe, Application of the Six Sigma concept in clinical laboratories: a review, Clin Chem Lab Med. 45 (6) (2007) 789–796.
[9] J.O. Westgard, S.A. Westgard, The quality of laboratory testing today: an assessment of sigma metrics for analytic quality using performance data from proficiency testing surveys and the CLIA criteria for acceptable performance, Am J Clin Pathol. 125 (3) (2006) 343–354.
[10] G. Cooper, N. DeJonge, S. Ehrmeyer, J. Yundt-Pacheco, R. Jansen, C. Ricos, et al., Collective opinion paper on findings of the 2010 convocation of experts on laboratory quality, Clin Chem Lab Med. 49 (5) (2011) 793–802.
[11] S.A. Westgard, H. Bayat, J.O. Westgard, A multi-test planning model for risk based statistical quality control strategies, Clin Chim Acta. 523 (2021) 216–223.
[12] J.O. Westgard, Q.C. Basic, Practices, 4th ed., Westgard QC, Madison, WI, 2016.
[13] J.M. Gras, Laboratory Quality Control and Patient Safety, De Gruyter, 2017.
[14] C.H. Schoenmakers, A.J. Naus, H.J. Vermeer, D. van Loon, G. Steen, Practical application of Sigma Metrics QC procedures in clinical chemistry, Clin Chem Lab Med. 49 (11) (2011) 1837–1843.
[15] A. Clerico, A. Padoan, M. Zaninotto, C. Passino, M. Plebani, Clinical relevance of biological variation of cardiac troponins, Clin Chem Lab Med. 59 (4) (2021) 641–652.
[16] WestgardQC. Desirable Specifications for Total Error, Imprecision, and Bias, derived from intra- and inter-individual biologic variation.
[17] R. El Sharkawy, S. Westgard, A.M. Awad, A.O.I. Ahmed, E.H. Iman, A. Gaballah, et al., Comparison between Sigma metrics in four accredited Egyptian medical laboratories in some biochemical tests: an initiative towards sigma calculation harmonization, Biochem Med (Zagreb). 28 (2) (2018), 020711.
[18] J. Afrifa, SethA Gyekye, WilliamKBA Owiredu, RichardKD Ephraim, S. Essien-Baidoo, S. Amoah, DavidL Simpong, AaronR Arthur, Application of sigma metrics for the assessment of quality control in clinical chemistry laboratory in Ghana: A pilot study, Niger Med J. 56 (1) (2015) 54.
[19] S. Cao, X. Qin, Application of Sigma metrics in assessing the clinical performance of verified versus non-verified reagents for routine biochemical analytes, Biochem Med (Zagreb). 28 (2) (2018), 020709.
[20] T. Fasano, J.L. Bedini, P.-A. Fle, M. Jlaiel, K. Hubbert, H. Datta, V. Chicha-Cattoir, H. Mansour, A. Mira, B.G. de la Presa, N. Rico, M.S. de Pedro, J. Diaz-Garzon, A. Buño Soto, L. Vecchia, K. Peoc'h, Multi-site performance evaluation and Sigma metrics of 20 assays on the Atellica chemistry and immunoassay analyzers, Clin Chem Lab Med. 58 (1) (2020) 59–68.
[21] Ganji SB, Revupalli S. Evaluation of Quality Assurance in a New Clinical Chemistry Laboratory by Six Sigma Metrics. Journal of Clinical and Diagnostic Research. 2019;13(3):BC04-BC7.
[22] P. Goel, G. Malik, S. Prasad, I. Rani, S. Manhas, K. Goel, Analysis of performance of clinical biochemistry laboratory using Sigma metrics and Quality Goal Index, Pract Lab Med. 23 (2021) e00195.
[23] Ö. Gülbahar, M. Kocabıyık, M.Z. Çıracı, C. Demirtaş, F. Uçar, N. Bayraktar, F. Akbıyık, M. Taşlıpınar, The use of six sigma methodology to evaluate the analytical performances of clinical chemistry analyzers, Turkish Journal of Biochemistry. 43 (1) (2018) 1–8.
[24] X. Guo, T. Zhang, X. Gao, P. Li, T. You, Q. Wu, J. Wu, F. Zhao, L. Xia, E. Xu, L. Qiu, X. Cheng, Sigma metrics for assessing the analytical quality of clinical chemistry assays: a comparison of two approaches: Electronic supplementary material available online for this article, Biochem Med (Zagreb). 28 (2) (2018), 020708.
[25] 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 (7) (2014) 973–980.
[26] B.V. Kumar, T. Mohan, Sigma metrics as a tool for evaluating the performance of internal quality control in a clinical chemistry laboratory, J Lab Physicians. 10 (2) (2018) 194–199.
[27] M. Lakshman, B.R. Reddy, P. Bhulaxmi, K. Malathi, M. Salma, S. Prakasham, Evaluation of sigma metrics in a Medical Biochemistry lab, Int J Biomedical Research. 6 (3) (2015) 164.
[28] R. Li, T. Wang, L. Gong, P. Peng, S. Yang, H. Zhao, P. Xiong, Comparative analysis of calculating sigma metrics by a trueness verification proficiency testing-based approach and an internal quality control data inter-laboratory comparison-based approach, J Clin Lab Anal. 33 (9) (2019), e22989.
[29] Y. Liu, Y. Cao, X. Liu, L. Wu, W. Cai, Evaluation of the analytical performance of endocrine analytes using sigma metrics, J Clin Lab Anal. 35 (1) (2021), e23581.
[30] X. Mao, J. Shao, B. Zhang, Y. Wang, Evaluating analytical quality in clinical biochemistry laboratory using Six Sigma, Biochem Med (Zagreb). 28 (2) (2018), 020904.
[31] S.K. Nanda, L. Ray, Quantitative application of sigma metrics in medical biochemistry, J Clin Diagn Res. 7 (12) (2013) 2689–2691.
[32] R. Nar, D.I. Emekli, The Evaluation of Analytical Performance of Immunoassay Tests by using Six-sigma Method, J Med Biochem. 36 (4) (2017) 301–308.
[33] S. Peng, J. Zhang, W. Zhou, W. Mao, Z. Han, Practical application of Westgard Sigma rules with run size in analytical biochemistry processes in clinical settings, J Clin Lab Anal. 35 (3) (2021), e23665.
[34] M.S. Shaikh, B. Moiz, Analytical performance evaluation of a high-volume hematology laboratory utilizing sigma metrics as standard of excellence, Int J Lab Hematol. 38 (2) (2016) 193–197.
[35] B. Singh, B. Goswami, V.K. Gupta, R. Chawla, V. Mallika, Application of sigma metrics for the assessment of quality assurance in clinical biochemistry laboratory in India: a pilot study, Indian J Clin Biochem. 26 (2) (2011) 131–135.
[36] M. Teshome, A. Worede, D. Asmelash, Total Clinical Chemistry Laboratory Errors and Evaluation of the Analytical Quality Control Using Sigma Metric for Routine Clinical Chemistry Tests, J Multidiscip Healthc. 14 (2021) 125–136.
[37] M.T. Tran, K. Hoang, R.F. Greaves, Practical application of biological variation and Sigma metrics quality models to evaluate 20 chemistry analytes on the Beckman Coulter AU680, Clin Biochem. 49 (16–17) (2016) 1259–1266.

[38] M. Verma, K. Dahiya, V.S. Ghalaut, V. Dhupper, Assessment of quality control system by sigma metrics and quality goal index ratio: A roadmap towards preparation for NABL, World J Methodol. 8 (3) (2018) 44–50.
[39] J. Xia, S.-F. Chen, F. Xu, Y.-L. Zhou, Quality specifications of routine clinical chemistry methods based on sigma metrics in performance evaluation, J Clin Lab Anal. 32 (3) (2018) e22284.
[40] Y. Zeng, H. He, K. Qin, M. Zhang, Z. An, H. Huang, Practical application of the sigma-metric run size nomogram for multistage bracketed statistical quality control analysis of eight enzymes, Clin Chim Acta. 492 (2019) 57–61.
[41] B. Zhou, Y. Wu, H. He, C. Li, L. Tan, Y. Cao, Practical application of Six Sigma management in analytical biochemistry processes in clinical settings, J Clin Lab Anal. 34 (1) (2020), e23126.
[42] CLSI. Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions, 4th Edition; CLSI guideline C24-ED4:2016 Wayne, PA Clinical and Laboratory Standards Institute; 2016.
[43] M. Feldhammer, M. Brown, J. Colby, J. Bryksin, B. Milstid, J.H. Nichols, A Survey of Sigma Metrics across Three Academic Medical Centers, J Appl Lab Med. 6 (5) (2021) 1264–1275.
[44] A.D. Ellis, A.R. Gross, J.R. Budd, W.G. Miller, Influence of reagent lots and multiple measuring systems on estimating the coefficient of variation from quality control data; implications for uncertainty estimation and interpretation of QC results, Clin Chem Lab Med. 58 (11) (2020) 1829–1835.
[45] R. Carboni-Huerta, K.V. Saenz-Flor, Sigma and Risk in the Quality Control Routine: Analysis in Chilean Clinical Laboratories, J Appl Lab Med. 7 (2) (2022) 456–466.
[46] 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, Clin Chem Lab Med. 53 (6) (2015).
[47] F. Braga, M. Panteghini, Performance specifications for measurement uncertainty of common biochemical measurands according to Milan models, Clin Chem Lab Med. 59 (8) (2021) 1362–1368.
[48] Aarsand A.K. F-CP, Webster C., Coskun A., Gonzales-Lao E., Diaz-Garzon J., Jonker N., Minchinela J., Simon M., Braga F., Perich C., Boned B., Roraas T., Marques-Garcia F., Carobene A., Aslan B., Barlett WA., Sandberg S. The EFLM Biological Variation Database [November 5, 2021]. Available from: https://biologicalvariation.eu/.
[49] A. Carobene, A.K. Aarsand, W.A. Bartlett, A. Coskun, J. Diaz-Garzon, P. Fernandez-Calle, E. Guerra, N. Jonker, M. Locatelli, M. Plebani, S. Sandberg, F. Ceriotti, The European Biological Variation Study (EuBIVAS): a summary report, Clin Chem Lab Med. 60 (4) (2022) 505–517.
[50] W.A. Bartlett, F. Braga, A. Carobene, A. Coşkun, R. Prusa, P. Fernandez-Calle, T. Røraas, N. Jonker, S. Sandberg, on behalf of the Biological Variati, A checklist for critical appraisal of studies of biological variation, Clin Chem Lab Med. 53 (6) (2015).
[51] B. Varela, G. Pacheco, Comprehensive evaluation of the internal and external quality control to redefine analytical quality goals, Biochem Med (Zagreb). 28 (2) (2018), 020710.
[52] J.O. Westgard, S.A. Westgard, Establishing Evidence-Based Statistical Quality Control Practices, Am J Clin Pathol. 151 (4) (2019) 364–370.
[53] WestgardQC. QC Frequency Calculator (Run Size) [January 4, 2022]. Available from: https://www.westgard.com/qc-frequency-calculator.htm.
[54] W.P. Oosterhuis, A. Coskun, Sigma metrics in laboratory medicine revisited: We are on the right road with the wrong map, Biochem Med (Zagreb). 28 (2) (2018), 020503.
[55] S. Westgard, H. Bayat, J.O. Westgard, Mistaken assumptions drive new Six Sigma model off the road, Biochem Med (Zagreb). 29 (1) (2019), 010903.
[56] A. Coskun, M. Serteser, M. Kilercik, F. Aksungar, I. Unsal, A new approach to calculating the Sigma Metric in clinical laboratories. Accred, Qual Assur. 20 (2) (2015) 147–152.
[57] A. Coskun, M. Serteser, I. Unsal, Sigma metric revisited: True known mistakes, Biochem Med (Zagreb). 29 (1) (2019), 010902.
[58] W.P. Oosterhuis, M.J. Severens, Performance specifications and six sigma theory: Clinical chemistry and industry compared, Clin Biochem. 57 (2018) 12–17.