Clinical Biochemistry 63 (2019) 106–112

Contents lists available at ScienceDirect

Clinical Biochemistry

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

A novel Sigma metric encompasses global multi-site performance of 18 assays on the Abbott Alinity system

Jennifer Taher[a,1], Jake Cosme[a,1], Brian A. Renley[b], David J. Daghfal[b], Paul M. Yip[a,c,*]

[a] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
[b] Diagnostics Division, Abbott Laboratories, Abbott Park, IL, USA
[c] Department of Clinical Biochemistry, University Health Network, Toronto, ON, Canada

ARTICLE INFO

Keywords:
Sigma metric
Chemistry
Immunoassay
Alinity
Architect

ABSTRACT

Objectives: The Abbott Alinity family of chemistry and immunoassay systems recently launched with early adopters contributing imprecision and bias data, which was consolidated to assess the performance of Alinity assays across multiple sites using the Sigma metric. Multi-site Sigma metrics were determined for 3 ion-selective electrodes, 12 photometric assays, and 3 immunoassays across 11 independent laboratory sites in 9 countries.
Methods: Total allowable error (TEa) goals followed a previously defined hierarchy that used CLIA as the primary goal. Bias was calculated against the Abbott ARCHITECT system using Passing-Bablok regression analysis using individual site data or pooled aggregate data. Sigma metrics were calculated as *(%TEa - |% bias|)/%CV*. For individual-site analysis, the Sigma metrics for each assay were compared using the individual-site and the pooled biases. For multi-site analysis, the average CV and the pooled bias were used to generate a Pooled Sigma metric encompassing the global performance for a given assay.
Results: A total of 97 individual-site and 18 Pooled Sigma metrics were calculated for available assays. Individual Sigma metrics varied across sites, with 90% of assays performing 4 Sigma or higher, and 17 of 18 Pooled Sigma metrics indicated performance greater than 4 Sigma. Sigma metrics were significantly improved in 16 assays when using pooled bias rather than individual-site bias.
Conclusions: This multi-center study applies a novel application of Sigma metrics to the first Alinity users and reveals analytical performance of greater than 4 Sigma for vast majority of assays. Laboratories with limited resources can leverage larger data sets for Pooled Sigma metric analysis, providing a tool to assess the consistency of analytical performance from multiple sites.

1. Introduction

Sigma metrics are a widely applied indicator of quality that has gained use in laboratory medicine to assess the performance of an analytical method. It is often used to indicate the degree of reliability of an assay for clinical use and the expected rate of error for an assay [1].

Quantitatively, the Sigma metric is a composite measure of the method's total allowable error (TEa), bias, and imprecision, whereby the source of each of these components will influence the final value. While TEa has been established through a hierarchy of sources [2], the preferred option is a derivation from outcome studies or biological variation. When these sources are not available or impractical for current technologies, the state-of-the-art goals are used, which is the most common. A previous report has shown that the choice for TEa goals is important in calculating an appropriate Sigma metric [3].

Imprecision data is derived often from in-house examinations using quality control (QC) material from defined process designs [4]. This would include multiple levels of QC materials measured in replicates in as little as 5 days, a month, or possibly even longer. Bias can be determined from different approaches. The CLSI guidelines describe bias determination using reference materials [4] or with patient specimens [5] depending on the users' perspective. Additionally, laboratories typically subscribe to proficiency testing (PT) or external quality assessment (EQA) programs as part of their regulatory requirements, where the bias of a given result can be determined relative to the peer group mean, all-methods mean, or values from a reference method.

In clinical laboratories, Sigma metrics have aided in the assessment of assay or instrument performance. The use of Sigma metrics can be used to implement more appropriate QC rules and practices [6]. Sigma metrics have also been used to evaluate assays using short-term

* Corresponding author at: Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Room B204, Toronto M4N 3M5, ON, Canada
E-mail address: paul.yip@utoronto.ca (P.M. Yip).
[1] co-first authors.

https://doi.org/10.1016/j.clinbiochem.2018.10.003
Received 25 May 2018; Received in revised form 3 October 2018; Accepted 8 October 2018
Available online 10 October 2018

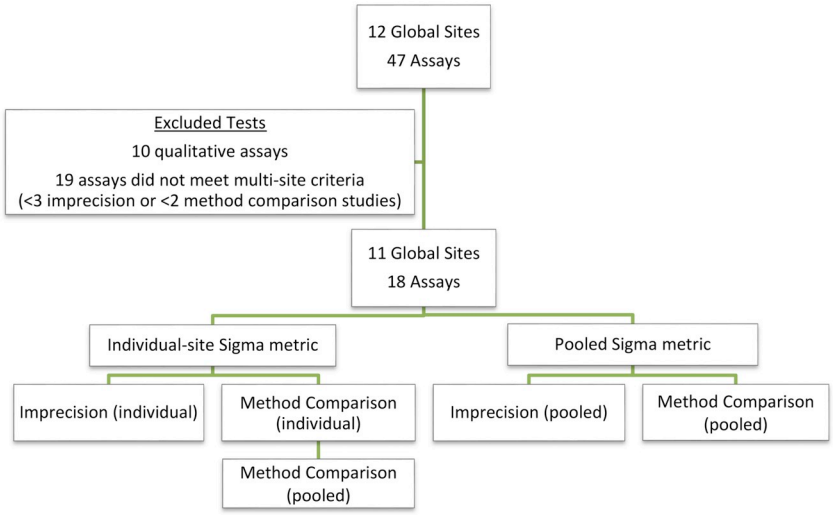

Fig. 1. Selection chart of the laboratory sites and assays tested, tests excluded, and Sigma metric calculations performed from individual or pooled imprecision and method comparison data. For individual-sites that generated imprecision and sufficient method comparison ($n \geq 40$), Sigma metric values were calculated using the individual-site bias from the individual method comparison, in addition to the Pooled Sigma metric for each assay.

imprecision and bias [7]. Lastly, instrument-specific QC rules have been implemented based on Sigma metrics along with long term monitoring of Sigma performance across an organizational network [8].

Evaluation of QC performance is based partly on the probability of rejecting an analytical run that contains a critical error, presenting as systematic or random error. The Sigma metric can be related to the rejection characteristics of a QC procedure to select appropriate control rules and number of control measurements for a given laboratory assay. However, this approach does not address the frequency with which QC should be performed. More recent CLSI documents have promoted the use of risk management strategies to design a laboratory's QC program and take into account QC frequency. This concept is highlighted in both CLSI EP23-A [9] and CLSI C24-A4 [10], although both provide limited data on implementation of the recommended practices [11]. The Sigma metric is a predictor of risk and can be used in the design of risk-based statistical QC procedures [12] or aid in the determination of QC frequency that limits risk to patient safety [13,14].

The Abbott Alinity family contains recently launched automated chemistry (Alinity c) and immunoassay (Alinity i) systems. Previously, the Alinity immunochemistry systems were evaluated for Sigma metrics using a large subset of assays that were tested within the manufacturer's facility [15]. The study was performed using an individual-site approach for Sigma metric calculation. In the current study, early users of the Alinity system voluntarily shared imprecision and method comparison data for this analysis. As a result, a large data set of real-world performance was generated for the Alinity system including both imprecision estimates and method comparison data against the current generation ARCHITECT platforms.

With the availability of imprecision and bias data from multiple international sites, the goal of this study was to assess the performance of the Alinity platform using Sigma metrics both from an individual site but also from a multiple-site perspective. The study also analyzes the components of method comparison studies before and after combining data from individual sites. This concept of a *Pooled Sigma metric* was shown to improve when calculated as a group, especially where a broad range of samples was lacking for individual method comparison studies.

2. Materials and methods

Assay imprecision and method comparison data was collected from 12 laboratory sites across 9 countries (Supplemental Table 1). The method verification data was generated from the Alinity i or Alinity ci system (Abbott Laboratories, Abbott Park, IL, USA) as part of a voluntary commercial program during laboratory implementation. The Alinity i is a fully automated immunoassay analyzer that utilizes chemiluminescent detection for the quantification of analytes. The Alinity ci is an integrated clinical chemistry and immunoassay analyzer that incorporates an additional module with photometric detection technology for the chemistry assays and also Integrated Chip Technology (ICT) for potentiometric detection of Na^+, K^+ and Cl^-.

Across the 12 sites, imprecision and method comparison studies were assessed using manufacturer's reagents for 47 assays based on CLSI guidelines [4]. All method comparisons were against the ARCHITECT system at each site. Inclusion criteria for multi-site Sigma analysis included: 1) quantitative test results for reporting purposes, and 2) assays that had ≥ 3 imprecision studies and ≥ 2 method comparison studies. A selection of 2 or more method comparison studies allowed for sufficient data to assess the effect of method comparison differences in Sigma calculations. A final total of 18 assays were included in the Sigma analysis and comprised of 3 ICT, 3 immunoassay, and 12 clinical chemistry assays (Supplemental Table 2). Elimination of 29 assays resulted in the exclusion of 1 site that did not have data on the selected assays, thereby leaving 11 laboratory sites in the final analysis (Fig. 1).

Multiple sources of TEa are available, with no international standard. However, a primary aim of the current manuscript is to assess the performance of the Alinity systems under multi-site, customer-based conditions, and benchmark to the previous data obtained internally by the manufacturer [15]. To compare results between controlled and real-world conditions on assay performance (Supplemental Table 4), we have followed the pre-selected TEa goals for each assay based on the previously defined hierarchy by Westgard [15]. The hierarchy uses CLIA goals as a primary choice and when values were not available or appropriate, the following sources were used in sequential order: College of American Pathologists (CAP), Ricos Biological Variation Database, Royal College of Pathologists in Australasia (RCPA) EQA program, Spanish minimum goals EQA program, and German RiliBÄK. Allowable limits for high sensitivity (HS) Troponin I TEa was not included in the previous study [15] and the TEa limit for this analyte was selected from Ricos Biological Variation Database (Table 1).

For each site, between-day imprecision data was generated for each assay according to CLSI EP15-A3 guidelines [4]. All sites performed 5-day or 20-day imprecision studies using multiple replicates per day for 2-3 levels of QC material, with the exception of two sites that carried out 20-day imprecision using 1 run per day. Sigma metric calculations for each assay were determined for individual sites and collectively as follows. For individual-site CV, the imprecision was taken from the QC concentration closest to a target level as previously defined [15]. The target level reflects a clinically relevant analyte concentration where a medical decision would be made. For pooled CV, the target level selected was based on the overall mean of the selected QC levels from the

Table 1
TEa goals and source. Allowable error shown in italics are adjusted from the CLIA acceptable performance into percentage values in proportion to the medical decision level [15].

Module	Assay	TEa	Source
ICT	Sodium	*4%*	CLIA
ICT	Potassium	*18%*	CLIA
ICT	Chloride	*5%*	CLIA
CC	ALP	30%	CLIA
CC	ALT	20%	CLIA
CC	Amylase	30%	CLIA
CC	AST	20%	CLIA
CC	Bilirubin, Total	20%	CLIA
CC	Calcium	*9.72%*	CLIA
CC	Total CO_2	25%	CAP 3SD approximate
CC	Glucose	10%	CLIA
CC	Magnesium	25%	CLIA
CC	Phosphorus	10.7%	CAP
CC	Protein, Total	10%	CLIA
CC	Urea	9%	CLIA
IA	β-hCG	30%	RiliBÄK
IA	HS Troponin I	27.91%	Ricos Desirable
IA	TSH	*23.7%*	Ricos Desirable

individual-site imprecision studies (Supplemental Table 3A). Where available, 20-day imprecision studies were chosen over 5-day imprecision studies. For Pooled Sigma metric calculations, a pooled total CV (referred to as *pooled CV*) was calculated from the average imprecision of individual sites (Equation 1) giving equal weighting to each site.

$$Pooled\ CV = \sqrt{\frac{(Total\ CV_{Site\#1})^2 + (Total\ CV_{Site\#2})^2 \ldots + (Total\ CV_{Site\#n})^2}{\#of\ Sites}} \quad (1)$$

Method comparison data was generated for each assay according to CLSI EP09-A3 guidelines [5]. All method comparison studies were performed against the corresponding assay on the Abbott ARCHITECT analyzer at the same site and analyzed using Passing-Bablok regression analysis. Residuals were calculated for each of the models, then inspected for normal distribution and centering around zero. For each assay, individual-site bias was calculated when there was a minimum of 40 method comparison samples. The bias (referred to as the *pooled bias*) was calculated from the pooled method comparison data across all sites. For sites that did not submit method comparison analysis, or completed it with less than 40 samples, only pooled bias was calculated. For Sigma metric calculations, bias at the target level was calculated by interpolation of the regression equation of the method comparison data.

Sites that generated both imprecision and sufficient method comparison data have two Sigma metric values using individual-site bias (equation 2) and across-site bias (equation 3). Pooled Sigma metrics were calculated using the pooled bias and pooled CV as shown in equation 4.

$$\frac{\%TEa - |\%bias|}{\%CV} \quad (2)$$

$$\frac{\%TEa - |\%pooled\ bias|}{\%CV} \quad (3)$$

$$\frac{\%TEa - |\%pooled\ bias|}{\%pooled\ CV} \quad (4)$$

3. Results

3.1. Multi-site Sigma metrics

From the initial 12 sites in this study, 18 assays were included in the final analysis. These 18 assays report quantitative test results and had ≥3 imprecision studies and ≥2 method comparison studies. Twenty-nine assays were excluded which did not meet the multi-site study criteria. Also, one site was excluded that did not perform any of the 18 candidate assays in its verification studies.

Initial Sigma metric analysis was performed using individual-site imprecision data and pooled method comparison where all method comparison data points were compiled into a single comparison study. In total, 97 Sigma metrics were calculated on an individual-site basis across the 18 assays (Supplemental Table 3b). A Sigma metric value cannot be calculated when imprecision is stated as an interval since it requires a finite value. As a result, two sites that obtained imprecision of less than 0.1% CV for potassium assays were excluded from Sigma metric calculation for lack of individual-site imprecision values.

A pooled total CV was calculated for each assay from imprecision data from all of the selected sites. Pooled CV values ranged from 1.0% to 6.5% across all testing platforms (Table 2). Pooled method comparison data against the predicate ARCHITECT platform showed bias at the target level and was typically less than half of the TEa goal, with pooled bias ranging from -2.2% to 2.5% for the 18 assays.

Table 2
Number of pooled imprecision and method comparison studies for Alinity ci assays and associated pooled CV, percent bias and Sigma metric values. (ICT, Integrated Circuit Technology assays; CC, Clinical Chemistry assays; IA, Immunoassay)

Module	Assay	Number of Studies		Level	Pooled %Bias	Pooled CV	Pooled Sigma
		Imprecision	Method Comparison				
ICT	Sodium	9	6	121 mmol/L	< 0.1%	1.0%	4.1
ICT	Potassium	7	6	2.7 mmol/L	< 0.1%	1.5%	11.8
ICT	Chloride	8	6	96 mmol/L	< 0.1%	1.0%	5.2
CC	ALP	8	4	151 U/L	-1.1%	2.7%	10.6
CC	ALT	4	3	28 U/L	-0.4%	3.3%	6.0
CC	Amylase	4	3	132 U/L	< 0.1%	1.1%	27.0
CC	AST	8	4	40 U/L	< 0.1%	2.5%	8.0
CC	Bilirubin, Total	3	3	48 µmol/L	< 0.1%	2.6%	7.6
CC	Calcium	6	6	2.45 mmol/L	1.0%	1.6%	5.3
CC	Total CO_2	3	2	21 mmol/L	< 0.1%	6.5%	3.9
CC	Glucose	6	5	6.6 mmol/L	-0.8%	1.6%	5.7
CC	Magnesium	4	2	0.77 mmol/L	0.7%	1.5%	16.7
CC	Phosphorus	3	2	1.46 mmol/L	-2.2%	1.6%	5.4
CC	Protein, Total	7	4	49 g/L	-2.0%	1.3%	6.2
CC	Urea Nitrogen	4	4	17.3 mmol/L	< 0.1%	1.6%	5.7
IA	β-hCG	3	4	26 IU/L	2.5%	6.2%	4.4
IA	HS Troponin I	6	5	19 ng/L	-0.9%	4.9%	5.5
IA	TSH	4	4	0.16 mIU/L	-1.1%	2.3%	10.0

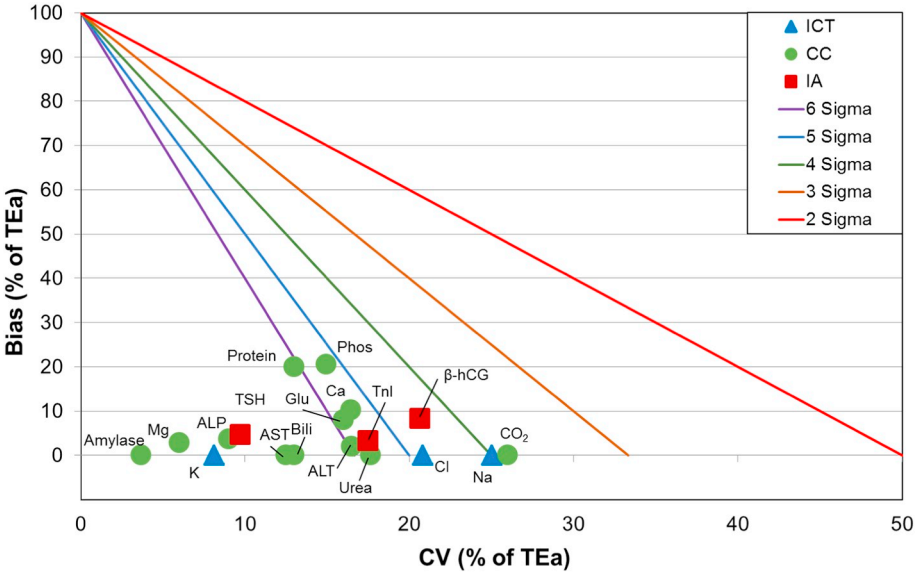

Fig. 2. Normalized Method Decision Chart for Alinity ci Assays, using Pooled Sigma metrics. CV and bias are normalized to the TEa, to separate the components of error for each assay's performance at the target level. (ICT, Integrated Circuit Technology assays; CC, Clinical Chemistry assays; IA, Immunoassay)

To assess the performance of each assay, we calculated a Pooled Sigma metric. At the medically relevant target level of each assay, 9 of 18 assays performed at 6 Sigma or better, 15 of 18 assays performed at 5 Sigma or better, and 17 of 18 assays performed at 4 Sigma or better. The one exception was total CO_2, which approached 4 Sigma with a metric of 3.9. This Pooled Sigma metric is shown as a normalized method decision chart (Fig. 2) where the imprecision and % bias of each assay is shown in relation to the TEa [16].

Comparison of the Sigma metrics derived from different sources of imprecision (individual-site or pooled) was also assessed. Individual-site Sigma metrics, using pooled bias data, varied around the Pooled Sigma metric (Fig. 3). For example, the Alinity calcium assay was used broadly by several sites, where 6 sites contributed to the pooled CV and 6 sites contributed to the pooled method comparison, such that variability was observed around the Pooled Sigma metric. Individual-site imprecision ranged from 0.9% to 2.4% CV at the target level, with an average pooled CV of 1.6%. Consolidation of method comparison data for calcium resulted in a pooled bias of 1.0% with individual-site biases ranging from 1.0 to 1.1%. Although the allowable error of 9.72% is one of the more stringent acceptability limits in the selected assays of the study, the individual sites performed with at least 3 Sigma (Range: 3.7 to 10.2) and together showed a Pooled Sigma metric of 5.3.

3.2. Differences in Sigma metric between single-site and multi-site analysis

In some cases we observed a significant difference in the resulting Sigma metric that depended on whether it was derived from the individual-site bias or pooled bias sources. For assays with a Sigma metric less than 6, there were 56 instances where an individual-site Sigma metric was calculated. When using pooled bias, in place of individual-site bias, Sigma values showed an improvement across most assays with an average increase of 0.52 in the Sigma metric (range -2.2 to 5.2).

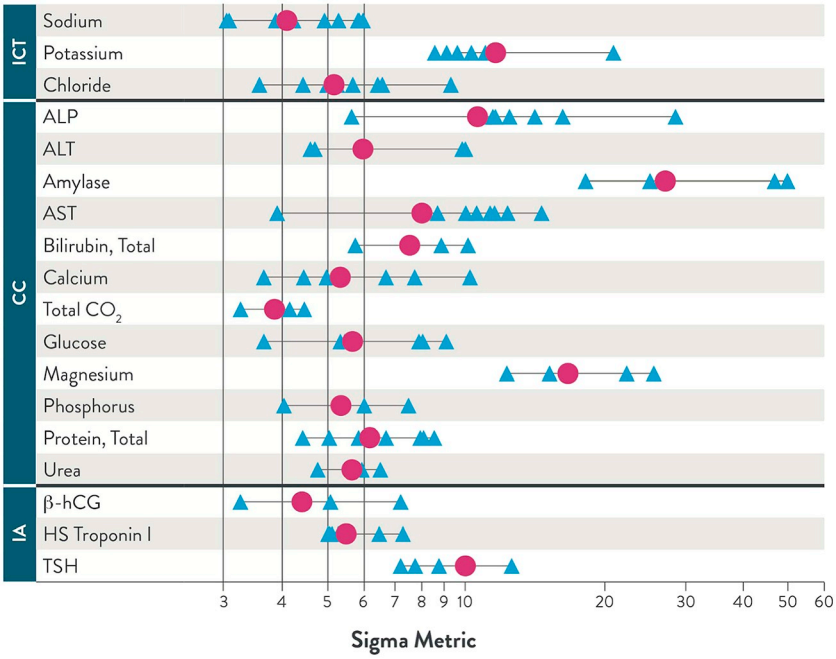

Fig. 3. Performance of 18 Alinity assays evaluated using Sigma metrics with individual and pooled imprecision data. Each assay's performance was evaluated using individual-site imprecision and pooled bias (blue), or pooled imprecision and pooled bias (red). Pooled imprecision and pooled bias represent a Pooled Sigma metric to reflect a composite metric across all selected sites. (ICT, Integrated Circuit Technology assays; CC, Clinical Chemistry assays; IA, Immunoassay)

Table 3
Investigation of Sigma level performance reclassification due to differences in the bias source

Assay	Site	Sigma metric		Main Source of Bias Difference	Method comparison factor contributing to bias difference
		Individual- site bias	Pooled bias		
Sodium	3	2.4	4.2	Constant	Minimal sample size (n = 40)
Sodium	6	4.2	5.3	Constant	Not determined
Sodium	4	5.2	5.8	Constant	Majority of points in a narrow range
Chloride	3	4.4	5.0	Constant and proportional	Minimal sample size (n = 40) and narrow range (44.2% of multi-site)
Chloride	4	5.2	6.5	Constant	Majority of points in a narrow range
AST	10	2.5	3.9	Constant and proportional	Narrow range (Range is 14.2% of multi-site)
ALT	11	4.0	4.7	Constant and proportional	Narrow range (Range is 20.6% of multi-site)
Calcium	10	4.6	5.0	Constant and proportional	Narrow range (Range is 42.7% of multi-site)
CO2	10	2.6	3.3	Constant	Narrow range (Range is 36.0% of multi-site)
Phosphorus	11	3.0	4.0	Proportional	Narrow range (Range is 26.3% of multi-site)
Total Protein	1	4.1	5.9	Proportional	Fewer points below 25th centile (8.9% of points)
Urea	6	4.3	5.8	Constant and Proportional	Minimal sample size (n = 40)
βhCG	8	2.7	3.3	Constant	Fewer points in lower range (n = 4 below target level)
HS Troponin I	8	4.4	5.0	Proportional	Heavy distribution near lower limit of AMR (88% of points below 25th centile)
HS Troponin I	4	3.3	5.3	Proportional	Fewer points in lower range

Twenty-seven assays improved with pooled bias, 10 assays decreased in Sigma, and 19 assays were unchanged. At Sigma values less than 6, a difference of 0.5 was considered significant since this would change the classification of the Sigma level performance. A 0.5 Sigma increase at a level below 5 Sigma reflects a larger improvement in quality than at 6 Sigma and higher (i.e. world-class performance). A change of one level of Sigma performance can impact the design of a QC program. Based on this, the pooled bias reclassified an individual-site Sigma performance for 15 assays in the data set as shown in Table 3.

We then examined the linear regression equations from the individual-site method comparison studies. As shown in Table 3, the differences at the individual site's target level were identified either to be constant bias (y-intercept), proportional bias (slope), or both. The regression equations of the method comparison studies are summarized in Supplemental Table 3b. By examination of the regression slope or intercept, the difference in bias (individual-site or pooled) was typically attributed to uneven distribution of values across the assay analytical measuring range (AMR), uneven distribution around the target level, or low number of samples for comparison (Table 3).

4. Discussion

The class or level of performance that is based on the Sigma metric following a method evaluation study has practical implications for an assay, including the design of QC programs [13]. Studies have mostly examined single-instrument Sigma metric analysis of assay performance for a given platform [15]. In this study, we have applied a novel approach to Sigma metrics following the implementation of the next generation Abbott Alinity analyzer across global sites in comparison with the predicate assays on the Abbott ARCHITECT in an independent manner.

Prior assessment of the Alinity system demonstrated that > 90% of the assays performed at 5 and 6 Sigma [15]. The strength of the current multi-site analysis is the assessment of the Alinity system under field conditions. In contrast to the previous study within the manufacturer's facility, the customer sites in this study did not perform extensive analytical method validation studies. However, our current study evaluated Sigma metrics using the same TEa and similar target levels to compare against previous benchmarks for the Alinity system. Unsurprisingly, our multi-site study found lower Sigma metric values than the previous internal study data. However, 6 Sigma level performance was sustained for 9 assays, and 13 of 17 assays remained at higher than 5 Sigma with the multi-site analysis (Supplemental Table 4). Given the previously high Sigma metric, these differences should not reflect a substantial change in performance. The remaining 4 assays were sodium, chloride, total CO_2, and β-HCG, that were mostly 1 Sigma class lower mainly due to higher estimates of imprecision.

Sigma metric analysis of the Alinity HS Troponin I assay is first reported here and has not been assessed previously. Due to the lack of practical goals available in the previously defined hierarchy, a TEa of 27.91% was selected based on Ricos Biological Variation Database. Among all 18 assays, HS Troponin I showed the least variability in Sigma metric across 6 sites (Fig. 3) with consistently tight CV values (3.7-5.4%) at the target level of 19 ng/L which is consistent with the low imprecision requirement to measure in the normal range [17]. Despite this, a Sigma level difference was observed in 2 sites that improved with a pooled bias rather than an individual-site bias (Table 3). These differences were attributed to low sample size and poor distribution across the AMR for the initial method comparison studies. Nonetheless the Pooled Sigma metric performance was greater than 5 Sigma for all sites.

In the setting of routine clinical testing, laboratory end-users should expect a similar level of performance when using the same platform and a common family of reagents. Deviations in an individual-site Sigma metric calculation can identify local problems such as instrumentation malfunction, reagent issues, sample handling, or potential operator-related errors. These factors influence performance that can be revealed during assay verification at the end-user site, which includes at a minimum, imprecision and method comparison studies. In the absence of any specific performance issues, statistical factors may contribute to poor method comparison analysis including small sample size and uneven distribution of values across the assay's AMR.

As outlined by CLSI guideline EP09-A3 [5], a proper method comparison would require 40 unmodified patient specimens covering the measuring interval of the two methods under evaluation [5], with levels spanning the low to high measuring interval. Furthermore high samples may be difficult to obtain. Through the use of a pooled method comparison, one should be able to obtain a broad distribution of samples by combining data across multiple sites. This approach provides an alternative to PT samples or spiked-in patient specimens, which have been

deemed as last-resort alternatives by the CLSI guideline EP09-A3 [5]. Similarly, concerns are minimized regarding matrix effects or analyte recovery that may reduce the commutability of the method comparison results as seen with some PT specimens [18].

By merging the method comparison data from individual sites for a given assay, a larger data set was achieved with additional data points that span a broader range across the AMR or improved the estimate at the target level. Also, this approach may better reflect an assay's real-world performance by end users. The combination of multiple method comparison studies to provide an improved bias determination was able to ameliorate some of the issues that an individual site may encounter. Our data suggests that the inclusion of multi-site data within the same instrument family or platform can allow for sufficient samples to extend the range of concentrations across the AMR or avoid having the regression equation skewed by a single high or low analyte specimen. Overall the pooled bias of the selected assays on the Alinity platform is well within the TEa and comparable to the current ARCHITECT platform, essentially allowing for equivalency of results in the transition to the newer instrumentation.

Macro-level analysis of Sigma metrics is emerging in clinical chemistry where it has been applied to PT survey data [19]. In this previous study, results from 5 different PT programs across Canada, USA, UK and Australia were used to assess performance of 29 chemistry assays. Imprecision was determined from the instrument group standard deviation. However, in the absence of a definitive reference method, bias was excluded from Sigma calculations. Sigma metrics varied significantly across manufacturer, instrument and assay method. While the study preceded the availability of the Alinity system, the Abbott ARCHITECT demonstrated the highest proportion of chemistry assays (78%) that performed at greater than 5 Sigma in comparison to several other automated chemistry analyzers [19].

Although the current study used the ARCHITECT as the reference analyzer, our multi-site Sigma metric analysis has several advantages over PT materials: 1) inclusion of bias in the Sigma metric calculation, 2) data was aggregated from method comparison studies using patient specimens as commutable materials, and 3) broad distribution of assay results that are not necessarily available in a convenient manner. While the assumption of zero bias has been often applied for practical reasons, a recent study illustrates the importance of bias in assessing performance characteristics to estimate the reliability of hemoglobin A1c results [20]. Through a robust method comparison of the ARCHITECT versus Alinity, our study included a pooled bias across multiple sites that was accounted in the Sigma metric calculation (Table 4). On the other hand, PT samples provide greater accessibility to varied analyte concentrations, procedural consistency across multiple sites and manufacturer methods. Thus, Sigma metric analysis using patient specimens and PT samples can be used to complement one another when comparing different platforms.

5. Limitations

We evaluated the analytical performance using Sigma metrics which followed previous criteria for TEa. No universally accepted TEa goals exist and thus harmonization is needed. Various recommendations are available for TEa [21,22] with ongoing debate regarding the selection of practical TEa goals. Even the most stringent criteria that are based on biological variation have been cited as too demanding for important analytes such as sodium and bicarbonate [3]. Although the TEa ultimately influences the Sigma metric, the relative rank order of an assay among different systems remains the same when comparing best to lowest performance. A second limitation of the study was the different protocols for imprecision studies carried out across sites, which may explain the Sigma metric differences between sites for a given assay. Similarly, our inclusion criteria limited the analysis to 18 assays and represented only one-third of those previously studied [15]. As the implementation of new Alinity systems increases, it may be possible to expand the voluntary program and the assays for inclusion in the Pooled Sigma analysis. Third, another limitation is the lack of comparison to a higher order reference method. Therefore the Sigma metrics only describe relative accuracy of the Alinity system to its predecessor. The Alinity and ARCHITECT systems showed minimal bias, which was likely due to similar reagent formulation and analytical parameters. Lastly, the assessment of assay performance occurred at a single concentration where imprecision and bias were assessed at the concentration of the control material nearest to the a medical decision level. The source of QC materials differed between sites and occasionally had disparate values. To address this issue, laboratories with imprecision data having concentrations dissimilar with the selected target level were excluded from the analysis. Importantly, assay performance at one concentration does not indicate similar performance at another, as shown by previous Sigma metric analyses where reported bias and imprecision of an assay varied across the AMR [23].

6. Conclusions

We show that 17 assays on the Alinity i and Alinity c systems have analytical performance of greater than 4 Sigma using a multi-site approach for combining imprecision and method verification data. Favorable Sigma metrics were a result of both low imprecision and bias, indicating good comparability between the Alinity and ARCHITECT systems across independent sites. The pooled method comparison data circumvents the limitations of single-site bias estimates. In doing so, we show the impact of different sources of bias to highlight factors contributing to the Sigma metric calculation. Our novel concept of the Pooled Sigma metric was used to compare variation in Sigma level performance across multiple sites. The Pooled Sigma metric could potentially be used as a tool by laboratories to assess inter-instrument analytical performance and consistency across multiple independent sites.

Acknowledgements

We would like to acknowledge Tricia Ravalico and Randy Schneider for their critical review of the manuscript. We also thank the contributions of: Manuel Doering, Elisabeth Kreutz, Juliane Fazio, Birgit Hartmeier, Jerome Motol, Clare Ford, Hayley Sharrod-Cole, Kevin Padmore, Chistine Chesters, Magdalena Krintus, F. Papa, Daniele Frattolillo, Jacques Mairesse, Ursula Koeller, Jose Luis Mirabel, and Carmen Zugazasalazer in data collection from participant sites.

Disclosure

Brian Renley and David Daghfal are employees of Abbott Laboratories. Jennifer Taher, Jake Cosme and Paul Yip received honoria and travel support from Abbott Diagnostics.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] N. Nikolac, M. Panteghini, E. Theodorsson, G.L. Salvagno, M. Miler, A.-M. Simundic, et al., How to assess the quality of your analytical method? Clinical Chemistry and Laboratory Medicine 53 (2015) 1707–1718.
[2] S. Sandberg, C.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, et al., Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine, Clinical Chemistry and Laboratory Medicine 53 (2015) 833–835.
[3] 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, Clinical Chemistry and Laboratory Medicine 52 (2014) 973–980.

[4] CLSI, User Verification of Precision and Estimation of Bias - Approved Guideline—Third Edition. CLSI document EP15-A3, Clinical and Laboratory Standards Institute, Wayne, PA, 2014.
[5] CLSI, Measurement Procedure Comparison and Bias Estimation Using Patient Samples - Approved Guideline—Third Edition. CLSI document EP09-A3, Clinical and Laboratory Standards Institute, Wayne, PA, 2013.
[6] C.H.H. Schoenmakers, A.J.M. Naus, H.J. Vermeer, D. van Loon, G. Steen, Practical application of Sigma Metrics QC procedures in clinical chemistry, Clinical Chemistry and Laboratory Medicine 49 (2011) 1837–1843.
[7] M. Ris, S. Božičević, V.R. Biljak, M.V. Lovrenčić, Analytical verification and quality assessment of the Tosoh HLC-723GX HbA1c analyzer, Practical Laboratory Medicine 7 (2017) 15–18.
[8] H.H. Harrison, J.B. Jones, Using Sigma Quality Control to Verify and Monitor Performance in a Multi-Instrument, Multisite Integrated Health Care Network, Clinics in Laboratory Medicine 37 (2017) 207–241.
[9] CLSI, Laboratory Quality Control Based on Risk Management - Approved Guideline. CLSI guideline EP23- A, Clinical and Laboratory Standards Institute, Wayne, PA, 2011.
[10] CLSI, Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions-4th Edition, Clinical and Laboratory Standards Institute, Wayne, PA., 2016.
[11] C.A. Parvin, What's new in laboratory statistical quality control guidance? The 4th edition of CLSI C24, statistical quality control for quantitative measurement procedures: principles and definitions, Journal of Applied Laboratory Medicine 1 (2017) 581–584.
[12] J.O. Westgard, S.A. Westgard, Six sigma quality management system and design of risk-based statistical quality control, Clinics in Laboratory Medicine 37 (2017) 85–96.
[13] M. Yago, S. Alcover, Selecting statistical procedures for quality control planning based on risk management, Clinical chemistry 62 (2016) 959–965.
[14] H. Bayat, S.A. Westgard, J.O. Westgard, Planning Risk-Based Statistical Quality Control Strategies: Graphical Tools to Support the New Clinical and Laboratory Standards Institute C24-Ed4 Guidance, The Journal of Applied Laboratory Medicine: An AACC Publication 2 (2017) 211–221.
[15] S. Westgard, V. Petrides, S. Schneider, M. Berman, J. Herzogenrath, A. Orzechowski, Assessing precision, bias and sigma-metrics of 53 measurands of the Alinity ci system, Clinical Biochemistry 50 (2017) 1216–1221.
[16] V.S. Smolcic, L. Bilic-Zulle, Normalized MEDx chart as a useful tool for evaluation of analytical quality achievements. A picture is worth a thousand words, Clinical Chemistry and Laboratory Medicine 51 (2013) e99–e101.
[17] P.O. Collinson, D. Gaze, S. Goodacre, The clinical and diagnostic performance characteristics of the high sensitivity Abbott cardiac troponin I assay, Clinical Biochemistry 48 (2015) 275–281.
[18] W.G. Miller, G.R.D. Jones, G.L. Horowitz, C. Weykamp, Proficiency testing/external quality assessment: current challenges and future directions, Clinical Chemistry 57 (2011) 1670–1680.
[19] S.A. Westgard, Utilizing global data to estimate analytical performance on the Sigma scale: A global comparative analysis of methods, instruments, and manufacturers through external quality assurance and proficiency testing programs, Clinical Biochemistry 49 (2016) 699–707.
[20] A. Woodworth, N. Korpi-Steiner, J.J. Miller, L.V. Rao, J. Yundt-Pacheco, L. Kuchipudi, et al., Utilization of assay performance characteristics to estimate hemoglobin A1c result reliability, Clinical Chemistry 60 (2014) 1073–1079.
[21] G.R.D. Jones, K. Sikaris, J. Gill, Allowable limits of performance for external quality assurance programs–an approach to application of the Stockholm criteria by the RCPA Quality Assurance Programs, Clinical Biochemist Reviews 33 (2012) 133.
[22] CLIA proficiency testing criteria for acceptable analytical performance - Federal Register Volume 57, Issue 40 (February 28, 1992), Office of the Federal Register, National Archives and Records Administration. 7002 - 7186.
[23] M.T.C. 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, Clinical Biochemistry 49 (2016) 1259–1266.