RESEARCH ARTICLE

WILEY

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

Runqing Li [ID] | Tengjiao Wang | Lijun Gong | Peng Peng | Song Yang | Haibin Zhao | Pan Xiong

Department of Laboratory Medicine, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

Correspondence
Runqing Li, Department of Laboratory Medicine, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, No. 168 Litang Road, Changping District, Beijing, 102218, China.
Email: lrqa00335@btch.edu.cn

Abstract

Introduction: Two methods were compared for evaluating the sigma metrics of clinical biochemistry tests using two different allowable total error (TEa) specifications.

Materials and methods: The imprecision (CV%) and bias (bias%) of 19 clinical biochemistry analytes were calculated using a trueness verification proficiency testing (TPT)-based approach and an internal quality control data inter-laboratory comparison (IQC)-based approach, respectively. Two sources of total allowable error (TEa), the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88) and the People's Republic of China Health Industry Standard (WS/T 403-2012), were used to calculate the sigma metrics (σ_{CLIA}, $\sigma_{WS/T}$). Sigma metrics were calculated to provide a single value for assessing the quality of each test based on a single concentration level.

Results: For both approaches, $\sigma_{CLIA} > \sigma_{WS/T}$ in 18 out of 19 assays. For the TPT-based approach, 16 assays showed $\sigma_{CLIA} > 3$, and 12 assays showed $\sigma_{WS/T} > 3$. For the IQC-based approach, 19 and 16 assays showed $\sigma_{CLIA} > 3$ and $\sigma_{WS/T} > 3$, respectively.

Conclusions: Both methods can be used as references for calculating sigma metrics and designing QC schedules in clinical laboratories. Sigma metrics should be evaluated comprehensively by different approaches.

KEYWORDS
allowable total error, internal quality control, proficiency testing, sigma metrics

1 | INTRODUCTION

Sigma (σ) metric was first introduced into clinical laboratories by David Nevalainen[1] in 2000. The practice of using the σ to improve clinical laboratory quality has been in place for nearly two decades.[2] Sigma metric has become a useful tool to monitor quality indicators,[1] to assess the analytical quality of assays,[3,4] to set quality control rules,[5-8] to describe assay analytical performance for external quality assessment participants,[9] and to help manufacturers choose product requirements.[10] Analytical quality of assays is quantitatively estimated as a sigma metric based on 3 parameters: allowable total error (TEa), bias, and imprecision. The TEa from various sources, such as the

Clinical Laboratory Improvement Amendments (CLIA), the College of American Pathologists (CAP), Reference Institute for Bioanalytics (Rilibak), the Royal College of Pathologists of Australasia (RCPA), and the China National Center for Clinical Laboratories (NCCL) external quality assessment goals, is associated with significantly different Sigma metrics for the same assay.[8,11] Thus, laboratories have to take into account which TEa specification is most appropriate initially.

Imprecision is usually presented as the standard deviation (SD) or coefficient of variation (CV). CV (SD) and bias often vary with the concentration of analytes, so the σ computed at different QC concentration levels could be quite different from one another.[12] Previous studies have reported that individual assays' biases were calculated by comparing their mean with the statistical target mean in external quality assessment (EQA) survey reports,[13,14] but the statistical target mean was derived from statistical results, which did not have measurement traceability. Additionally, if the concentration of IQC materials differed significantly from that of EQA samples, the σ calculation may not represent the optimal quality of the analytes. On the other hand, the target means in trueness verification proficiency testing was determined by the reference methods. Therefore, the bias and imprecision from this approach were more convincing.[12] Laboratories using commercial quality control could upload the IQC data to the inter-laboratory comparison database and compare these data with the target mean of the peer group to calculate bias and CV. The internal quality control data inter-laboratory comparison-based approach has been shown to be a convenient and reliable method among clinical laboratories. Xinqi Cheng et al[15] introduced two approaches to compute bias and CV for σ evaluation, using a proficiency testing (PT) or an internal quality control (IQC)-based approach. Herein, we take similar approaches.[1] For the trueness verification proficiency testing (TPT)-based approach, bias was calculated by comparing the measured mean with the target mean derived from the measured results of the reference measurement procedure provided by the China National Center for Clinical Laboratories (NCCL), and the CV was calculated by testing the analytes five times daily for 3 days for each TPT sample according to the instructions from the Chinese NCCL[2]; for the internal quality control data inter-laboratory comparison (IQC)-based approach, bias was calculated by comparing the measured mean with the target mean of the global peer group, and CV was calculated with the IQC results. Based on both approaches, the bias and CV of the analyte could be acquired simultaneously at the same concentration, rather than being synthetically calculated at different concentrations. We introduced this study to compare both approaches for bias and CV calculations and to determine the effects of these approaches on σ evaluation at two TEa specifications.

2 | MATERIALS AND METHODS

2.1 | Materials

Nineteen assays were tested using the manufacturer's original reagents and calibrators on the Siemens Advia 2400 system (Siemens Healthcare Diagnostics Inc) in Beijing Tsinghua Changgung Hospital. The reference method/material traceability of the assays is shown in Table 1. The trueness verification proficiency testing (TPT) samples (including Metabolites and Total Protein/Electrolyte/Enzymes/Lipids trueness verification proficiency testing samples) for nineteen analytes were provided by China National Center for Clinical Laboratories (NCCL) in 2018. The nineteen analytes of Metabolites and Total Protein/Electrolyte/Enzymes/Lipids TPT samples were GLU, UN, CREA, UA, TP/Na, K, Ca, Mg, Cl/ALT, AST, GGT, LDH, CK, AMY, ALP/TC, and TG. The internal quality controls were the Bio-Rad (Bio-Rad Laboratories, Inc) Liquid Assayed Multiqual controls (Lot No. 45792/45793).

2.2 | Methods

2.2.1 | Sample preparation

All the TPT samples from the Chinese NCCL were prepared from pooled fresh human serum. Each level, with 3 separate aliquots, was frozen and stored at −70°C until shipment on dry ice to our laboratory. The samples were stored in the freezer at −70°C as soon as they were received until use. When testing, the samples were placed at room temperature for 20 minutes and then gently mixed until completely dissolved and assayed in our laboratory within 4 hours.

2.2.2 | Bias calculation

For the TPT-based approach, the target value of each level of nineteen analytes of TPT samples was assigned by the reference laboratories organized by the Chinese NCCL, following the recommended reference measurement procedures.[16-18] The aliquot of each level was measured on 3 specific days, and each aliquot was tested under repeatability condition five times referring to the procedure of CLSI EP15-A2.[19] Thus, a total of 15 results were obtained for each concentration level. The average of the 15 results was considered as the laboratory-tested value. All the results were reported to the online EQA platform developed by the Chinese NCCL (www.nccl.org.cn), and a summary report of the evaluation results can also be acquired on the online EQA platform.[16,18] Bias was determined as (our mean − target mean)/target mean × 100%.

For the IQC-based approach, all IQC data of nineteen analytes were included in the Bio-Rad global comparison program and the data were collected in the clinical laboratory of Beijing Tsinghua Changgung Hospital from July 1, 2018, to December 30, 2018. The monthly (December 2018) mean, with SD and CV, was calculated. Bias was calculated as (our mean − mean of peer group)/ mean of peer group × 100%.[15]

2.2.3 | Imprecision evaluation

Imprecision in this study was determined for each level of 19 analytes of TPT samples using the 15 results according to the formula defined in CLSI guideline EP15-A2[19] for the TPT-based approach.[18]

Bio-Rad Liquid Assayed Multiqual controls were assayed daily for the nineteen analytes and calculated with the IQC results for the IQC-based approach.

TABLE 1 Methods, units, traceability of the analytes on the Siemens Advia 2400 analyzer and number of laboratories included in each instrument peer group for the trueness verification proficiency testing (TPT) program of the Chinese NCCL, and the number of laboratories included in the Bio-Rad IQC global peer group reports

Analytes[a]	GLU	UN	CREA	UA	TP	Na	K	Ca	Mg	Cl
Methods	Hexokinase	Urease, UV	Enzymatic	Uricase, colorimetric	Biuret	ISE indirect	ISE indirect	Arsenazo III	Xylidyl blue	ISE indirect
Units	mmol/L	mmol/L	μmol/L	μmol/L	g/L	mmol/L	mmol/L	mmol/L	mmol/L	mmol/L
Reference method traceability	NA	CDC-Reference method	IFCC-Reference method	CDC candidate Reference method (uricase)	Biuret Reference method	CDC Flame photometry reference method	CDC Flame photometry reference method	Inductively coupled atomic emission method	Atomic absorption	NIST coulometric reference
Reference Material Traceability	NIST SRM 965a	NIST SRM 912a, NIST SRM 909b	IRMM/IFCC-4	NIST SRM 913a, NIST SRM 909	NIST SRM 927	NIST SRM 909b	NIST SRM 909b	NIST SRM 915, NIST SRM 90	NIST SRM 929	Method NIST SRM 919
Number of laboratories for NCCL TPT	241	241	276	246	318	320	320	318	292	319
Number of laboratories for Bio-Rad IQC	35	17	13	40	40	50	50	29	44	49
Number of IQC data points	1695	507	653	2096	2190	2538	2529	1086	2052	2404

Analytes[a]	ALT	AST	GGT	LDH	CK	AMY	ALP	TC	TG
Methods	UV without P5P	UV without P5P	G-glutamyl-carboxy-nitroanilide	Lactate to pyruvate	NAC activated	G7 PNP, Blocked	PNPP, AMP Buffer, IFCC-Reference	Cholesterol oxidase	Enzymatic, GPO-PAP
Units	U/L	U/L	U/L	U/L	U/L	U/L	U/L	mmol/L	mmol/L
Reference Method Traceability	IFCC-Reference Method at 37°C	IFCC-Reference Method at 37°C	IFCC-Reference Method at 37°C	IFCC-Reference Method at 37°C	IFCC-Reference Method	IFCC-Reference Method	IFCC-Reference Method	CDC-Reference Method (modified Abell-Kendall)	Reference Method
Reference Material Traceability	IRMM/IFCC-454	NA	IRMM/IFCC-452	IRMM/IFCC-453	IRMM/IFCC-455	IRMM/IFCC-456	NA	NIST SRM 909b	NIST SRM 909b
Number of laboratories for NCCL TPT	316	313	383	379	374	338	380	362	324
Number of laboratories for Bio-Rad IQC	21	20	18	13	10	29	32	5	5
Number of IQC data points	765	760	577	331	208	1744	1865	64	66

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AMY, amylase; AST, aspartate aminotransferase; Ca, calcium; CK, creatine kinase; Cl, chloride; CREA, creatinine; GGT, gamma-glutamyltransferase; GLU, glucose; IQC, internal quality control; K, potassium; LDH, lactate dehydrogenase; LDL, low density lipoprotein cholesterol; Mg, magnesium; NA, none available; Na, sodium; NCCL, China National Center for Clinical Laboratories, Peking, China; TC, total cholesterol; TG, triglycerides; TP, total protein; TPT, trueness verification proficiency testing; UA, uric acid; UN, urea nitrogen.

[a]All analytes were tested on the Siemens Advia 2400 analyzer using manufacturer's original reagents.

TABLE 2 Sigma metrics computed by the trueness verification proficiency testing (TPT)-based approach and the internal quality control data inter-laboratory comparison (IQC)-based approach

Item	Sample number	Mean	Target	Bias (%)	CV (%)	TEa$_{CLIA}$ (% [absolute value])	TEa$_{ws/t}$ (%)	σ_{CLIA}	$\sigma_{ws/t}$
GLU	201911	3.66	3.70	−1.08	0.55	10 (0.33)	7.00	16.22	10.76
	201912	9.24	9.43	−2.01	0.87	10 (0.33)	7.00	9.18	5.74
	45792	6.19	6.22	−0.49	1.41	10 (0.33)	7.00	6.74	4.62
	45793	19.04	19.35	−1.63	1.18	10 (0.33)	7.00	7.09	4.55
UN	201911	5.06	4.85	4.33	1.58	9 (0.71)	8.00	2.96[b]	2.32[b]
	201912[c]	13.94	13.68	1.90	0.57	9 (0.71)	8.00	12.46	10.70
	45792[c]	14.02	14.22	−1.35	3.61	9 (0.71)	8.00	4.01	3.48
	45793	24.51	25.07	−2.24	1.97	9 (0.71)	8.00	3.43	2.92[b]
CREA	201911	65.20	66.41	−1.82	0.80	15 (26.5)	12.00	16.48	12.73
	201912	728.60	703.38	3.59	0.60	15 (26.5)	12.00	19.02	14.02
	45792	159.02	161.67	−1.64	1.86	15 (26.5)	12.00	7.18	5.57
	45793	569.52	571.09	−0.28	1.47	15 (26.5)	12.00	10.01	7.97
UA	201911	247.00	258.30	−4.37	3.00	17	12.00	3.54	2.54[b]
	201912	519.00	519.30	−0.06	1.00	17	12.00	14.94	11.94
	45792	366.99	371.93	−1.33	1.45	17	12.00	10.81	7.36
	45793	570.12	578.20	−1.40	1.17	17	12.00	13.33	9.06
TP	201911[c]	57.00	57.48	−0.84	1.60	10	6.00	5.73	3.23
	201912	76.50	79.66	−3.97	1.60	10	6.00	3.77	1.27[b]
	45792[c]	55.83	54.51	2.43	1.90	10	6.00	3.98	1.88[b]
	45793	71.83	70.40	2.03	1.70	10	6.00	4.69	2.34[b]
Na	201811	127.30	126.30	0.79	0.50	(4)	4.00	5.00[a]	6.42[a]
	201812[c]	140.80	140.36	0.31	0.40	(4)	4.00	7.12[a]	9.23[a]
	45792[c]	141.81	141.89	−0.06	0.40	(4)	4.00	6.88[a]	9.85[a]
	45793	159.95	159.56	0.24	0.54	(4)	4.00	4.15[a]	6.96[a]
K	201811	5.82	5.72	1.84	0.52	(0.5)	6.00	13.17[a]	8.00[a]
	201812	5.97	5.93	0.62	0.34	(0.5)	6.00	23.15[a]	15.82[a]
	45792	4.20	4.18	0.47	0.67	(0.5)	6.00	16.00[a]	8.25[a]
	45793	7.77	7.70	0.91	0.82	(0.5)	6.00	7.17[a]	6.21[a]
Ca	201811	2.10	2.08	0.82	2.38	(0.25)	5.00	4.65[a]	1.76[a,b]
	201812[c]	2.34	2.33	0.65	2.14	(0.25)	5.00	4.69[a]	2.03[a,b]
	45792[c]	2.47	2.50	−1.39	1.66	(0.25)	5.00	5.49[a]	2.17[a,b]
	45793	3.17	3.20	−1.00	1.55	(0.25)	5.00	4.39[a]	2.58[a,b]
Mg	201811	0.77	0.80	−3.99	3.90	25	15.00	5.39	2.82[b]
	201812[c]	1.05	1.07	−1.96	1.90	25	15.00	12.13	6.86
	45792[c]	1.11	1.09	2.12	1.98	25	15.00	11.56	6.51
	45793	1.75	1.74	0.96	1.82	25	15.00	13.21	7.71
Cl	201811[c]	95.20	95.19	0.01	0.40	5	4.00	12.48	9.98
	201812	107.10	105.24	1.77	0.30	5	4.00	10.77	7.43
	45792[c]	99.46	99.24	0.22	1.05	5	4.00	4.55	3.60
	45793	120.55	120.97	−0.34	1.19	5	4.00	3.92	3.08

(Continues)

TABLE 2 (Continued)

Item	Sample number	Mean	Target	Bias (%)	CV (%)	TEa$_{CLIA}$ (% [absolute value])	TEa$_{ws/t}$ (%)	σ_{CLIA}	$\sigma_{ws/t}$
ALT	201801	35.70	37.70	−5.31	5.30	20	16.00	2.77[b]	2.02[b]
	201802	62.90	67.40	−6.68	2.20	20	16.00	6.05	4.24
	201803[c]	99.40	109.30	−9.06	1.70	20	16.00	6.44	4.08
	201804	153.60	169.40	−9.33	1.00	20	16.00	10.67	6.67
	201805	191.60	205.30	−6.67	1.00	20	16.00	13.33	9.33
	45792[c]	93.17	95.55	−2.48	2.48	20	16.00	7.06	5.45
	45793	200.51	207.21	−3.24	1.71	20	16.00	9.80	7.46
AST	201801	36.70	33.90	8.26	2.20	20	15.00	5.34	3.06
	201802	73.00	70.10	4.14	1.00	20	15.00	15.86	10.86
	201803[c]	108.40	105.20	3.04	1.30	20	15.00	13.05	9.20
	201804	173.50	167.40	3.64	0.60	20	15.00	27.27	18.93
	201805	202.20	199.70	1.25	0.80	20	15.00	23.44	17.19
	45792[c]	111.36	114.25	−2.53	1.59	20	15.00	10.99	7.84
	45793	272.14	279.18	−2.52	1.43	20	15.00	12.22	8.73
GGT	201801	44.50	49.00	−9.18	2.20	15	11.00	2.65[b]	0.83[b]
	201802	74.00	77.32	−4.29	1.40	15	11.00	7.65	4.79
	201803[c]	147.70	150.67	−1.97	0.60	15	11.00	21.72	15.05
	201804	194.00	194.55	−0.28	0.60	15	11.00	24.53	17.87
	201805	239.90	243.32	−1.41	0.80	15	11.00	16.99	11.99
	45792	85.37	87.23	−2.13	1.71	15	11.00	7.53	5.19
	45793[c]	133.32	137.50	−3.04	1.52	15	11.00	7.87	5.24
LDH	201801	155.30	157.16	−1.18	0.80	20	11.00	23.53	12.28
	201802	196.90	197.55	−0.33	0.60	20	11.00	32.78	17.78
	201803	291.90	292.56	−0.23	0.70	20	11.00	28.24	15.39
	201804[c]	411.90	421.00	−2.16	0.40	20	11.00	44.60	22.10
	201805	517.20	526.98	−1.86	0.70	20	11.00	25.91	13.06
	45792	176.88	177.43	−0.31	1.47	20	11.00	13.39	7.27
	45793[c]	427.66	431.08	−0.79	1.09	20	11.00	17.62	9.37
CK	201801	139.30	139.76	−0.33	0.90	30	15.00	32.97	16.30
	201802	351.70	345.19	1.89	0.60	30	15.00	46.85	21.85
	201803[c]	659.30	632.56	4.23	0.50	30	15.00	51.54	21.54
	201804	864.40	842.35	2.62	0.50	30	15.00	54.76	24.76
	201805	1147.20	112.84	3.09	0.70	30	15.00	38.44	17.01
	45792	254.45	257.41	−1.15	1.44	30	15.00	20.03	9.62
	45793[c]	629.34	645.79	−2.55	2.52	30	15.00	10.89	4.94
AMY	201801[c]	124.00	118.00	5.08	0.90	30	15.00	27.69	11.02
	201802	253.70	240.31	5.57	0.80	30	15.00	30.54	11.79
	201803	436.80	409.93	6.55	0.40	30	15.00	58.63	21.13
	201804	611.20	575.36	6.23	0.40	30	15.00	59.43	21.93
	201805	722.10	676.74	6.70	0.50	30	15.00	46.60	16.60
	45792[c]	133.96	134.37	−0.31	1.81	30	18.00	16.40	8.12
	45793	285.64	289.69	−1.40	1.65	30	18.00	17.33	8.24

TABLE 2 (Continued)

Item	Sample number	Mean	Target	Bias (%)	CV (%)	TEa$_{CLIA}$ (% [absolute value])	TEa$_{ws/t}$ (%)	σ_{CLIA}	$\sigma_{ws/t}$
ALP	201801	105.00	94.73	10.84	0.80	30	18.00	23.95	8.95
	201802	260.90	241.75	7.92	0.60	30	18.00	36.80	16.80
	201803	454.80	402.41	13.02	0.60	30	18.00	28.30	8.30
	201804	541.60	487.27	11.15	0.40	30	18.00	47.13	17.13
	201805	635.90	580.86	9.48	0.70	30	18.00	29.31	12.17
	45792	138.54	141.77	−2.28	3.77	30	18.00	7.35	4.17
	45793	284.20	291.80	−2.60	2.40	30	18.00	11.42	6.42
TC	201811	5.78	5.63	2.72	1.21	10	9.00	6.02	5.19
	201812	3.72	3.61	2.96	1.08	10	9.00	6.52	5.59
	45792	4.32	4.21	2.51	1.41	10	9.00	5.31	4.60
	45793	7.72	7.69	0.32	1.10	10	9.00	8.80	7.89
TG	201811	1.91	1.84	4.09	1.05	25	14.00	19.91	9.44
	201812[c]	1.30	1.23	5.69	1.54	25	14.00	12.54	5.40
	45792[c]	1.54	1.49	3.50	2.06	25	14.00	10.44	5.10
	45793	4.92	4.81	2.17	1.10	25	14.00	20.75	10.75

Abbreviations: ALP, alkaline phosphatase, U/L; ALT, alanine aminotransferase, U/L; AMY, amylase, U/L; AST, aspartate aminotransferase, U/L; Ca, calcium, mmol/L; CK, creatine kinase, U/L; Cl, chloride, mmol/L; CREA, creatinine, μmol/L; CV, coefficient of variation; GGT, gamma-glutamyltransferase, U/L; GLU, glucose, mmol/L; K, potassium, mmol/L; LDH, lactate dehydrogenase, U/L; Mg, magnesium, mmol/L; Na, sodium, mmol/L; TC, total cholesterol, mmol/L; TEaCLIA, allowable total error derived from USA Clinical Laboratory Improvement Amendments of 1988 (CLIA '88); TEaWS/T, allowable total error derived from the People's Republic of China Health Industry Standard (WS/T 403-2012); TG, triglycerides, mmol/L; TP, total protein, g/L; UA, uric acid, μmol/L; UN, urea nitrogen, mmol/L.

[a]σ is calculated by using absolute bias as the allowable total error (TEa), not marked indicates using percentage bias to calculate the σ.

[b]σ < 3.

[c]σCLIA and σWS/T were compared at similar concentration levels for each assay using TPT-based method and IQC-based method.

2.2.4 | Sigma calculation

The two TEa requirements, USA Clinical Laboratory Improvement Amendments of 1988—CLIA '88 (TEa$_{CLIA}$) and People's Republic of China Health Industry Standard—WS/T 403-2012 (TEa$_{WS/T}$), were selected to calculate σ for each assay, using the equation σ = (TEa − bias)/ SD (for concentration units) and σ = (TEa% − bias%)/ CV% (for percentage units). Both TEa requirements of the 19 assays are listed in Table 2. Excel 2016 software (Microsoft Corporation) was used for data analysis and graphing.

3 | RESULTS

3.1 | Sigma metrics from TPT-based approach

Sigma metrics of 19 assays computed by the TPT-based approach and IQC-based approach using two TEa requirements (σ$_{CLIA}$ and σ$_{WS/T}$) are listed in Table 2. The TEa$_{CLIA}$ used absolute bias for K, Na, and Ca assays, whereas TEa$_{WS/T}$ used percentage bias and was stricter than TEa$_{CLIA}$. Among the 19 assays by a TPT-based approach, 16 assays showed σ$_{CLIA}$ > 3, 10 assays (GLU, CREA, K, Cl, LDH, CK, AMY, ALP, TC, and TG) showed σ$_{CLIA}$ > 6, and 3 assays (UN, ALT, and GGT) showed σ$_{CLIA}$ < 3. On the other hand, 12 assays showed σ$_{WS/T}$ > 3, 8 assays (CREA, Na, K, Cl, LDH, CK, AMY, and

ALP) showed σ$_{WS/T}$ > 6, and 6 assays (UN, UA, Ca, Mg, ALT, and GGT) showed σ$_{WS/T}$ < 3. σ$_{CLIA}$ > σ$_{WS/T}$ in 18 out of 19 assays (all except Na).

3.2 | Sigma metrics from IQC-based approach

Sigma metrics of 19 assays computed by the IQC-based approach using two TEa requirements (σ$_{CLIA,}$ σ$_{WS/T}$) are shown in Table 2. For the IQC-based approach, 19 and 16 assays showed σ$_{CLIA}$ > 3 (GLU, CREA, UA, K, Mg, Cl, ALT, AST, GGT, LDH, CK, AMY, ALP, and TG exceeded 6σ levels) and σ$_{WS/T}$ > 3 (UA, Na, K, Mg, Cl, AST, LDH, AMY, ALP, TC, and TG exceeded 6σ levels), respectively. Similar to the TPT-based approach, σ$_{CLIA}$ > σ$_{WS/T}$ for all assays except Na.

3.3 | Comparative analysis of sigma metrics between both methods

Nine analytes (GLU, CREA, K, Cl, LDH, CK, AMY, ALP, and TG) had a σ$_{CLIA}$ above 6, and 6 analytes (Na, K, Cl, LDH, AMY, and ALP) had a σ$_{WS/T}$ above 6 from both methods, which is actually effectively in agreement.

As shown in Figure 1, the σ$_{CLIA}$ and σ$_{WS/T}$ derived from both methods for nineteen analytes were significantly different. For instance, the σ$_{CLIA}$ for the TPT-based approach vs the IQC-based approach of UN at similar concentration levels was 12.46 (201912) versus 4.01 (45792), respectively, and similarly, the σ$_{WS/T}$ was 10.70 (201912) vs

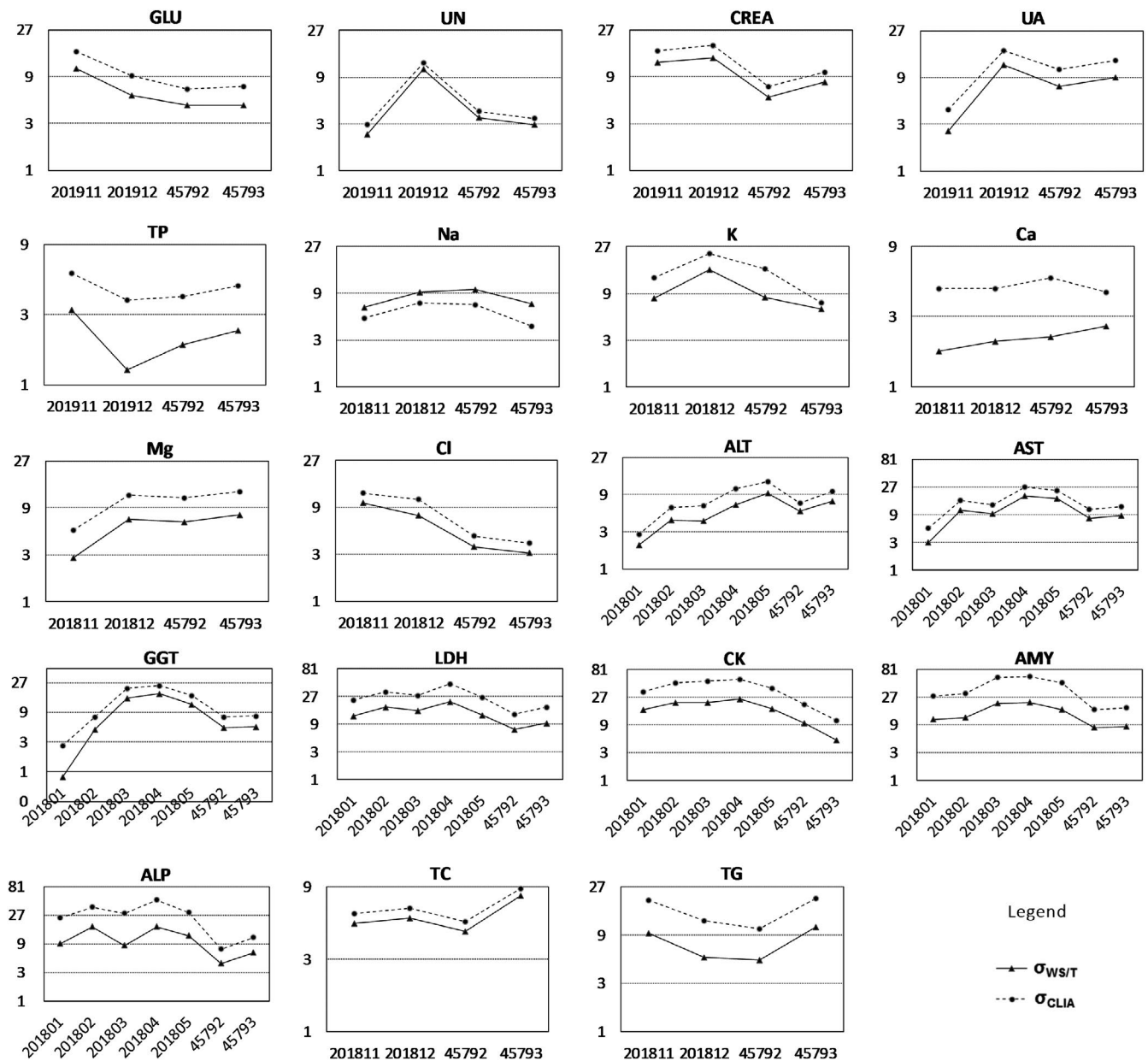**FIGURE 1** Comparison of σ_{CLIA} and $\sigma_{WS/T}$ calculated using the trueness verification proficiency testing (TPT)-based approach and the internal quality control data inter-laboratory comparison (IQC)-based approach for the same analyte. Note: solid lines are used for σ_{CLIA}, and dashed lines are used for $\sigma_{WS/T}$. 201801-201805, 201811-201812, and 201911-201912 represent the lot numbers of trueness verification proficiency testing materials, and 45792 and 45793 represent the lot number of Bio-Rad biochemistry quality control materials. ALP, alkaline phosphatase, U/L; ALT, alanine aminotransferase, U/L; AMY, amylase, U/L; AST, aspartate aminotransferase, U/L; Ca, calcium, mmol/L; CK, creatine kinase, U/L; Cl, chloride, mmol/L; CREA, creatinine, μmol/L; GGT, gamma-glutamyltransferase, U/L; GLU, glucose, mmol/L; K, potassium, mmol/L; LDH, lactate dehydrogenase, U/L; Mg, magnesium, mmol/L; Na, sodium, mmol/L; TC, total cholesterol, mmol/L; TG, triglycerides, mmol/L; TP, total protein, g/L; UA, uric acid, μmol/L; UN, urea nitrogen, mmol/L

3.48 (45792), respectively. The σ_{CLIA} and $\sigma_{WS/T}$ comparisons at similar concentration levels for both methods are shown in Table 2.

4 | DISCUSSION

Six Sigma and analytical sigma metrics are a widely accepted measure of assessing method quality, optimizing QC procedures, and processing improvement.[2,7] The sigma metrics computed at different concentration levels can be quite different from one another. Another challenge in calculating sigma metrics is how to acquire the appropriate TEa, bias, and CV (SD). Therefore, this study was introduced to explore the comparative analysis of sigma metrics computed by both methods.

There are different TEa quality requirements for routine chemistry,[20] and the optimal TEa quality requirements should be established

depending on the conditions and requirements of the individual laboratory. At present, the first choice of TEa selection in China is the WS/T 403-2012, which is the standard of the Ministry of Health published analytical quality specifications for routine clinical biochemistry in 2012. If the standard is too strict, other requirements, such as the CLIA'88, the College of American Pathologists (CAP), guidelines of the German medical association for the quality assurance of laboratory medical examinations (RiliBÄK), and The Royal College of Pathologists of Australasia quality assurance programs (RCPA), could also be used.[2,20] If the TEa is too strict, the laboratory should assess the suitability of the TEa for clinical use or patient care and determine if that would allow a larger TEa choice. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) published a recommended hierarchy for choosing an appropriate TEa whereby setting-required performance specifications are based on 3 possible models: clinical outcomes, biological variabilities, and state-of-the-art.[21] It provides a feasible model for standardizing protocols and is being introduced to encourage all laboratories to use the same quality goals for their sigma metrics benchmarks.[22] Because $TEa_{WS/T}$ was more stringent than TEa_{CLIA} for the most analytes, the $\sigma_{WS/T}$ was significantly lower than that of σ_{CLIA} in 18 out of 19 assays (all except Na). Therefore, the limitation of selecting an appropriate TEa requirement for σ calculation should be considered.

Bias is an estimate of systematic measurement error.[23] Assessing bias can be challenging, and bias evaluation can significantly impact the sigma metrics. Bias is always evaluated using the following methods.[8] (a) The optimal method is to compare results obtained from fresh human specimens using the measurement procedure and a reference measurement procedure. In this study, we used fresh frozen human serum samples provided by trueness verification proficiency testing program from the Chinese NCCL. The target means of nineteen analytes were determined using the corresponding reference measurement procedure from the Chinese NCCL. Samples derived from human serum could avoid the matrix effect, which was caused by the interaction of processed material and the measurement procedure and may suggest that erroneous results are being generated when in fact the results are acceptable.[24] The use of fresh frozen human specimens could avoid the matrix-related bias.[25] (b) Another way to assess relative bias is by comparing laboratory results with the statistical mean of the peer group using the same instrument and method from inter-laboratory QC data or proficiency testing (PT)/external quality assessment (EQA) reports. Because the target means in PT/EQA programs were derived from statistical results without measurement traceability,[15,26,27] those using this approach should be aware of the possible limitations, including statistical methods used to generate the data and the number of laboratories that participate.[23]

Imprecision is typically expressed as an SD or CV. The imprecision estimation for sigma metrics should be based on results from a sufficient duration of time to adequately represent the types of influences, such as a periodic recalibration, changes in bottles of reagents, changes in lots of reagents or calibrators, and maintenance procedures. We used six months of IQC data to obtain relatively stable results. By contrast, the TPT-based approach may lead to a lower CV and an overestimated σ value because of the short-term assessment.[15]

We further studied the QC rule design based on the sigma metrics of nineteen analytes designed based on sigma metrics by both methods using two TEa standards (data not shown in this study). The differences in QC rules, QC numbers, and QC run sizes as calculated using both approaches and two TEa standards were analyzed. The QC setting of nineteen analytes was compared test by test, that of seven analytes were the same and twelve analytes were different. QC rules based on $TEa_{ws/t}$ were more stringent than that based on TEa_{CLIA}. We are working on the adjustment of the individualized QC plan of the analyte based on its sigma metric evaluation.

One weakness of this study is that the imprecision evaluation following the Chinese NCCL instruction by the TPT approach (analyze five replicate samples at each of different concentrations daily for 3 days to obtain 15 results) was not conducted strictly according to the method described in the CLSI guideline EP15-A2 document (analyze one run per day with 3 replicate samples at each of different concentrations daily for 5 days to obtain 15 results) and CLSI guideline EP5-A3 (analyze two replicate samples at each of different concentrations daily for 20 days to obtain 40 results).[28] The TPT approach only lasted 3 days, which may have underestimated the imprecision.[18] The numbers of laboratories for the Bio-Rad IQC method from peer group comparison ranged between 5 and 50, and there may be increased imprecision due to the small number of laboratories. The concentration of analytes of the IQC materials differed significantly from that of the TPT samples, which may contribute to the differences in the calculation of σ based on both methods. Moreover, the research results may encourage the application of different methods for sigma evaluation, without the cost of high maintenance of commercial QC materials.

In conclusion, a combination of the TPT-based and IQC-based approach using NCCL and CLIA TEa goals may be adopted as the useful approaches for sigma metrics evaluation in clinical laboratories. When applying sigma metrics for Six Sigma management or quality control, sigma metrics should be evaluated comprehensively by different approaches.

CONFLICT OF INTEREST

None declared.

ORCID

Runqing Li https://orcid.org/0000-0003-3139-2871

REFERENCES

1. Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. *Arch Pathol Lab Med.* 2000;124:516-519.

2. Westgard S, Bayat H, Westgard JO. Analytical sigma metrics: a review of six sigma implementation tools for medical laboratories. *Biochem Med (Zagreb).* 2018;28:020502.

3. Maesa J, Fernández-riejos P, Sánchez-mora C, Toro-crespo M, González-rodriguez C. Application of six sigma model to evaluate the analytical quality of four HbA1c analyzers. *Clin Lab.* 2017;63(1):79-83.

4. Westgard S, Petrides V, Schneider S, Berman M, Herzogenrath J, Orzechowski A. Assessing precision, bias and sigma-metrics of 53 measurands of the Alinity ci system. *Clin Biochem.* 2017;50:1216-1221.

5. Bayat H. Selecting multi-rule quality control procedures based on patient risk. *Clin Chem Lab Med.* 2017;55:1702-1708.

6. Westgard JO, Bayat H, Westgard SA. Planning risk-based SQC schedules for bracketed operation of continuous production analyzers. *Clin Chem.* 2018;64:289-296.

7. Parvin CA. Planning statistical quality control to minimize patient risk: it's about time. *Clin Chem.* 2018;64:249-250.

8. Westgard S, Bayat H, Westgard JO. Special issue on six sigma metrics - experiences and recommendations. *Biochem Med (Zagreb).* 2018;28:020301.

9. Thelen M, Jansen R, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. *Clin Chem Lab Med.* 2017;55:1509-1516.

10. Petrides V, Schneider S. Using sigma metrics to establish analytical product performance requirements and optimize analytical performance of an in vitro diagnostic assay using a theoretical total PSA assay as an example. *Biochem Med (Zagreb).* 2018;28:020903.

11. Xia J, Chen SF, Xu F, Zhou YL. Quality specifications of routine clinical chemistry methods based on sigma metrics in performance evaluation. *J Clin Lab Anal.* 2018;32(3):e22284.

12. Li Runqing GL, Wang Tengjiao ZD, Yang Song ZH, Hu Xiaorui ZZ, Xiuying Z. The application of sigma metrics quality models in quality management of clinical biochemistry laboratory. *Chin J Lab Med.* 2017;40:727-732.

13. Shaikh MS, Moiz B. Analytical performance evaluation of a high-volume hematology laboratory utilizing sigma metrics as standard of excellence. *Int J Lab Hematol.* 2016;38:193-197.

14. Tran MT, Hoang K, Greaves RF. Practical application of biological variation and sigma metrics quality models to evaluate 20 chemistry analytes on the Beckman Coulter AU680. *Clin Biochem.* 2016;49:1259-1266.

15. Guo X, Zhang T, Gao X, et al. 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).* 2018;28:020708.

16. Wang Y, Wang J, Zhao H, et al. Assessment of enzyme measurement procedures in China through a trueness verification program. *Clin Chim Acta.* 2016;461:98-102.

17. Ge M, Zhao H, Yan Y, et al. Performance of electrolyte measurements assessed by a trueness verification program. *Clin Chem Lab Med.* 2016;54:1319-1327.

18. Kang F, Zhang C, Wang W, Wang Z. sigma metric analysis for performance of creatinine with fresh frozen serum. *Scand J Clin Lab Invest.* 2016;76:40-44.

19. CLSI. *User Verification of Performance for Precision and Trueness; Approved Guideline-Second Edition. CLSI Document EP15-A2.* Wayne, PA: Clinical and Laboratory Standards Institute; 2005.

20. Hens K, Berth M, Armbruster D, Westgard S. Sigma metrics used to assess analytical quality of clinical chemistry assays: importance of the allowable total error (TEa) target. *Clin Chem Lab Med.* 2014;52:973-980.

21. Sandberg S, Fraser CG, Horvath AR, et al. 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.* 2015;53:833-835.

22. El Sharkawy R, Westgard S, Awad AM, 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).* 2018;28(2):020711.

23. CLSI. *Statistical Quality Control for Quantitative Measurement Procedures; Principles and Definitions.4th edn. CLSI guideline C24.* Wayne, PA: Clinical and Laboratory Standards Institute; 2016.

24. CLSI. *Evaluation of Matrix Effects; Approved Guideline-Second Edition.* Wayne, PA: Clinical and Laboratory Standards Institute; 2005.

25. Miller WG, Myers GL. Commutability still matters. *Clin Chem.* 2013;59:1291-1293.

26. Friedecky B, Kratochvila J, Budina M. Why do different EQA schemes have apparently different limits of acceptability. *Clin Chem Lab Med.* 2011;49:743-745.

27. Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem.* 2011;57:1670-1680.

28. CLSI. *User Verification of Precision and Estimation of Bias; Approved Guideline-3rd edn. CLSI Document EP15-A3.* Wayne, PA: Clinical and Laboratory Standards Institute; 2014.