

RESEARCH ARTICLE

Assessment of patient based real-time quality control on comparative assays for common clinical analytes

Yide Lu¹ | Fan Yang¹ | Dongmei Wen² | Kaifeng Shi¹ | Zhichao Gu¹ | Qiuya Lu¹ | Xuefeng Wang^{1,3} | Danfeng Dong¹ 

¹Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

²Shanghai Senxu Medical Technology Corporation, Shanghai, China

³Faculty of Medical Laboratory Science, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Correspondence

Danfeng Dong and Xuefeng Wang, Department of Laboratory Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, No 197, Ruijin Er Road, Shanghai, China.
Emails: ddf40688@rjh.com.cn; wxf63@shsmu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81902117; the Shanghai "Rising Stars of Medical Talents" Youth Development Program-Clinical Laboratory Practitioner Program, Grant/Award Number: None

Abstract

Background: It is critical for laboratories to conduct multianalyzer comparisons as a part of daily routine work to strengthen the quality management of test systems. Here, we explored the application of patient-based real-time quality controls (PBRTQCs) on comparative assays to monitor the consistency among clinical laboratories.

Methods: The present study included 11 commonly tested analytes that were detected using three analyzers. PBRTQC procedures were set up with exponentially weighted moving average (EWMA) algorithms and evaluated using the AI-MA artificial intelligence platform. Comparative assays were carried out on serum samples, and patient data were collected. Patients were divided into total patient (TP), inpatient (IP), and outpatient (OP) groups.

Results: Optimal PBRTQC protocols were evaluated and selected with appropriate truncation limits and smoothing factors. Generally, similar comparative assay performance was achieved using both the EWMA and median methods. Good consistency between the results from patients' data and serum samples was obtained, and unacceptable bias was detected for alkaline phosphatase (ALP) and gamma-glutamyl transferase (GGT) when using analyzer C. Categorizing patients' data and applying specific groups for comparative assays could significantly improve the performance of PBRTQCs. When monitoring the inter- and intraanalyzer stability on a daily basis, EWMA was superior in detecting very small quality-related changes with lower false-positive alarms.

Conclusions: We found that PBRTQCs have the potential to efficiently assess multi-analyzer comparability. Laboratories should be aware of population variations concerning both analytes and analyzers to build more suitable PBRTQC protocols.

KEYWORDS

comparative assays, exponentially weighted moving average method, patient-based real-time quality control, quality management

Yide Lu and Fan Yang are contributed equally to this work

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Clinical Laboratory Analysis* published by Wiley Periodicals LLC.

1 | INTRODUCTION

At clinical laboratories, there is usually a great demand for the testing of biochemical samples. These tests may require the simultaneous use of multiple different models or brands of biochemical analyzers. The International Organization for Standardization (ISO) 15,189 standard specifies that a comparability scheme should be established to ensure the consistency of results. In routine practice, comparison assays with fresh serum samples¹ and daily quality control (QC) samples, which are not commutable as patient samples with a material matrix,²⁻⁴ are recommended. However, this system is conducted at certain intervals rather than continuously, which increases potential quality risks during the analytical process. Patient data have potential advantages in aiding the implementation of laboratory management,^{5,6} particularly following the revolution of information technology. Nevertheless, the consideration of applying patient data in multianalyzer comparisons is still underexplored.

The International Federation of Clinical Chemistry (IFCC) and the International Council for Clinical Chemistry and Analytical Quality in Laboratory Medicine proposed that patient-based real-time quality control (PBRTQC) is a valuable complementary quality control method, having the advantages of lower cost, absence of matrix effects, continuous real-time monitoring, and higher sensitivity to preanalytical errors.⁶ Various arithmetic procedures have been applied for PBRTQC, including Bull's algorithm,⁷ the moving average (MA) method,⁸ and the exponentially weighted moving average (EWMA) method.⁹ To evaluate and monitor the comparability and stability of clinical tests across manufacturers and laboratories, the median of patient data is most widely used.¹⁰ However, this requires a large quantity of samples, and the performance is greatly affected when the amount of data is relatively small. In the EWMA algorithm, a weighting coefficient is introduced to smooth the dataset by adjusting the sensitivity of bias based on the training data.¹¹ By using this method, which is reported to be less affected by sample size, small deviations in the process can be detected.

In this study, PBRTQC protocols using the EWMA algorithm were set up using the professional artificial intelligence software AI-MA. Then, we conducted comparison assays using the EWMA algorithm and the median method and evaluated the performance by comparing the results from serum samples. Overall, we attempted to establish a more convenient, economical, and efficient method for interanalyzer consistency monitoring in clinical laboratories.

2 | MATERIALS AND METHODS

2.1 | Analyzer and analytes

Three Beckman Coulter AU5800 series clinical chemistry analyzers (Beckman Coulter, United States) were used in the clinical laboratory at Ruijin Hospital, Shanghai Jiaotong University, School of Medicine,

namely, AU5821_1, AU5821_2, and AU5800_3, which were marked as analyzers A, B, and C, respectively, in the present study. Eleven common analytes, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), creatine (CREA), uric acid (UA), potassium (K), sodium (NA), chloride (CL), calcium (CA), and inorganic phosphorus (PHOS), were studied using the three analyzers. In total, approximately 5000 samples were measured every day. Analyzers A and B were randomly assigned to run outpatient and inpatient samples, which were equivalent to approximately 2000 tests per day. Analyzer C was mainly used for urgent samples, averaging 500 and 200 tests for inpatient and outpatient samples, respectively. Three levels of internal quality controls (IQC) were measured at the beginning of the analytical batch in the morning, and clinical sample testing was performed only when IQCs were in control. All of the IQC results were recorded and exported from the Beckman information system. Events that may have affected the analytical results, such as reagent lot changes, detection light calibrations, and reagent calibrations, were also recorded.

2.2 | Simulating and optimizing PBRTQC protocols based on the AI-MA platform

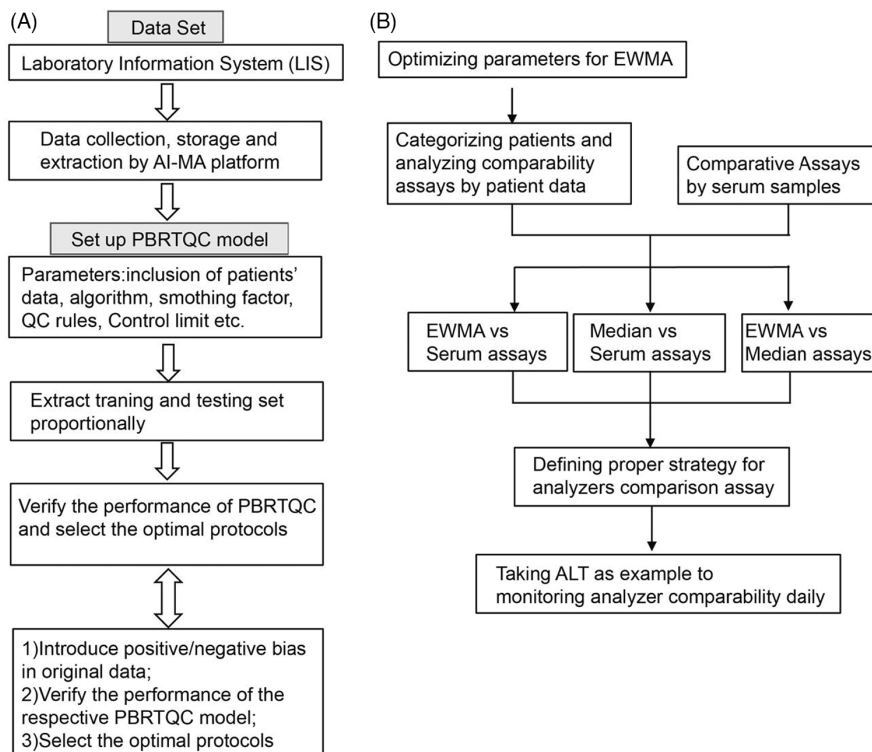
The process of establishing optimal PBRTQC protocols in the AI-MA platform is illustrated in Figure 1A. Briefly, raw data in the laboratory information system (LIS) containing measurand information were directly collected, extracted, and processed by using the AI-MA software version 1.0, which was developed by SENXU MEDICAL Corporation (Shanghai, China). By implementing multiple parameters for data investigation, a quality risk prediction model for PBRTQC was built through deep machine learning with the random forest method constructed from decision tree algorithms. The parameters, including the measurands, inclusion/exclusion of sample sources, specimen types, truncation limits, time windows, and smoothing factors, were set. The datasets for training and testing were proportionally extracted, and the PBRTQC model was established with multiple parameters. The EWMA algorithm is given as $EWMA_t = (1 - \lambda)EWMA_{t-1} + \lambda \times x_t$,¹² where $EWMA_t$ and x_t are the estimated and actual values of the t point measurement, respectively, and λ is the smoothing factor.

To obtain a stable dataset and avoid the deleterious effects of extreme values, we set truncation limits (TLs) based on biological variations (BVs) in the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Biological Variation Database.¹³ The BVs of a measurand contained within-subject variations (CV_I) and between-subject variations (CV_G). The upper and lower TLs of the measurands were derived from the following formulas:

$$RCV_G = 2.76 \times \sqrt{CV_G^2 + CV_I^2} \quad (1)$$

$$TLs = Means \pm RCV_G \quad (2)$$

FIGURE 1 Flow chart of (A) PBRTQC procedure and (B) comparative study design. Abbreviation: EWMA, exponentially weighted moving average; LIS, laboratory information system; PBRTQC, patient-based real-time quality control



Here, Means represents the average value of the test results from all patients within the previous 6 months. If the amount of data in each group per day did not exceed 10, the data were excluded.

The control limits of the AI-MA EWMA chart were set based on the traditional concept of the Levey–Jennings graph using $\bar{x} \pm 1, 2$ and $3SD$ and were modified by the introduction of quality objectives. \bar{x} and SD represent the mean and standard deviation of the EWMA results accumulated in the previous 6 months, respectively. EWMA QC rules were applied based on the 10_x and 4_{1s} "Westgard Rules" and machine learning of the control charts.

To verify the performance and select optimal protocols, we randomly selected a confirmed unbiased dataset and introduced positive/negative allowed total errors (TE_a)¹⁴ for approximately 200–300 data points. Under different parameters, the data were resubmitted and calculated. The probability of error detection (Ped) and false-positive rate (FPR) was calculated based on the quality control chart. The optimal EWMA protocols were determined if $Ped > 90\%$ and $FPR < 5\%$.

3 | COMPARATIVE ASSAYS BY SERUM SAMPLES AND PATIENT DATA FROM EWMA OR MEDIANS

The study design of the comparative assay is illustrated in Figure 1B. As the gold standard for interinstrument comparisons, measurement procedures were performed based on the Clinical and Laboratory Standards Institute (CLSI) EP9-A3 guidelines.¹⁵ Serum samples were collected and prepared according to an updated protocol based on Clinical and Laboratory Standard Institution (CLSI) C37-A.¹⁶ The

analyte concentrations were more evenly distributed within the measurement range and covered the medical determination levels (MDLs) (Table S1), as suggested at <https://www.westgard.com/decision.htm>. At least 40 fresh samples for each analyte, 8 per day for 5 consecutive days, were included. Single outlying results were identified and removed by the extreme linearized deviation (ESD) method. Extra samples were added to reach a total of 40 samples if necessary. The Passing-Bablok analysis model using Medcalc software version 19.3 was used to obtain regression equations. The assessment of bias was determined by substituting medical determination levels into the equations and setting the acceptance limits as $\frac{1}{2}$ allowed total errors (TE_a).¹⁷

During the period of serum sample comparisons, patient data from the three analyzers were collected using the median and the PBRTQC method under optimal parameter settings. Raw data of patient specimens were divided into three groups, total patient (TP), inpatient (IP) and outpatient (OP) groups, according to the source of the patients. Pairwise comparisons were performed, that is, B vs. A, C vs. A, and C vs. B. Interanalyzer differences were assessed by calculating the delta percentage difference of two compared results.

To further investigate the application of patient data in real-time monitoring interanalyzer comparability, the cumulative results of the three instruments were collected daily using the EWMA and median methods, and events that may have affected test results, including changes in reagent lots and bottles, the calibration of lights and kits, and the maintenance and malfunction of analyzers, were recorded. Intraanalyzer bias was assessed by comparing the daily EWMA or median value to its target value accumulated over the previous 6 months. Interanalyzer differences were calculated as $(\text{maximum-minimum})/\text{average} \times 100\%$.¹⁰ The acceptable bias limits of both the

intra- and interanalyzers were set to $\frac{1}{2} TE_a$. Inter- and intraanalyzer bias events were recorded. If interanalyzer deviation on a given day was due to bias caused by a certain instrument, the two events were considered the same; otherwise, they were calculated as two separate events. A true bias alarm suggested by patient data was determined in the presence of a corresponding quality event; otherwise, it was regarded as a false alarm.

4 | RESULTS

4.1 | Performance of the PBRTQC algorithms

Taking serum K as an example, two batches of positive and negative bias were introduced into the original data (Figure 2). The number of biases detected by the PBRTQC model on the AI-MA platform was recorded with various weighting coefficients (Table S2). Using the truncation limit helped to exclude extreme values and provided a more stable dataset for process monitoring. It was presumed that a better performance of PBRTQC could be achieved using this dataset than the dataset without truncation (Figure 2). After performing simulations with different smoothing factors, we chose a λ value of 0.03 for the optimal PBRTQC model, and a Ped of 91.54% and FRP of 0% were obtained (Table 1). Moreover, the number of patient results affected before error detection (NPed) was investigated under different EWMA parameters. The optimal PBRTQC model generated a smaller NPed (Table 1, Figure 2), indicating a better performance in detecting erroneous results in a timely manner. Accordingly, the parameters for other analytes were obtained and are listed in Table S3.

4.2 | Test comparability by serum samples

We conducted serum sample comparison testing from August 16, 2021, to August 20, 2021. Forty samples of each analyte were separately measured in the three analyzers without any outlying results. Regression analysis revealed that all of the analytes had good correlation coefficients ($r^2 > 0.99$) among the analyzers. Then, we assessed the bias in medical determination by pairwise comparisons to determine whether the deviation was acceptable. Table S1 provides a detailed overview of the comparative assay results by serum samples. In general, good comparability was achieved for analytes ALT, AST, CREA, UA, K, NA, CL, CA, and PHOS, as no unacceptable levels of bias were detected at either level of concentration for any of the manufacturers. However, for analytes ALP and GGT, significant bias was observed when using analyzer C, as increased test results occurred for these analytes at lower concentrations.

4.3 | Test comparability using patient data

Accordingly, we conducted comparison assays using patient data. Notably, Figure S1 shows that the test number for each analyte in

analyzer C was distinct from the other two analyzers, as there was a smaller number of OP patients. To investigate whether the results were affected by the patients' source, we analyzed the dataset used with analyzer A, which had an equivalent number for the IP and OP groups, under the respective optimal PBRTQC models. The accumulated EWMA values are listed in Table S4. For analytes ALP, GGT, UA, K, and CA, the differences between patients' sources were found to exceed the applied limits. Concerning distinct sample numbers and sources, all analytes were categorized to evaluate the performance of patient data in the analyzer comparison. The pairwise comparative analysis results by EWMA were in good accordance with those by the median method in each group (Table 2). For analytes with good comparability by serum assays, such as AST, K, NA, CL, CA, and PHOS, the assessment of pairwise bias was also performed with desirable limits derived from the EWMA and median methods in each group. For measurements of analytes ALT and UA, the serological comparison was excellent, but the results were falsely lower using analyzer C on the OP group for ALT and on the TP group for UA. Discrepancies when applying the data derived from the OP group were not detected for analytes GGT and ALP, for which we observed considerable positive bias when using analyzer C.

4.4 | Application of patient data in the surveillance of analyzer comparisons

To assess the application of the PBRTQC models in real-time analyzer comparisons, the ALT analyte data from the TP and IP groups were studied based on the aforementioned results. During the monitoring period, IQCs at the three concentrations were all in controls (Figure S2). The intra- and interanalyzer deviations by day were plotted with their respective symbols and corresponding dates (Figure 3). Overall, 3 and 4 potential abnormal events were recorded when using EWMA on the TP and IP groups, among which 66.67% (2/3) and 100% (4/4) were identified as true, respectively. In contrast, 6 and 10 out-of-range bias events were recorded for the TP and IP groups, respectively, when using the median method; however, only 50% of them were considered true (Table 3). Generally, the method using the median value had a significantly higher false-positive rate than the EWMA method. Categorizing patients when using EWMA protocols can clearly improve the accuracy of unexpected bias detection by increasing the detection rate and decreasing the false-positive rate.

5 | DISCUSSION

In this study, we aimed to establish optimal PBRTQC models for commonly requested test samples and assess their potential role in comparability assays in daily work. Compared with the synchronous serological methods, the results showed that PBRTQCs could be applied as an efficient tool for monitoring test comparability and stability in laboratories.

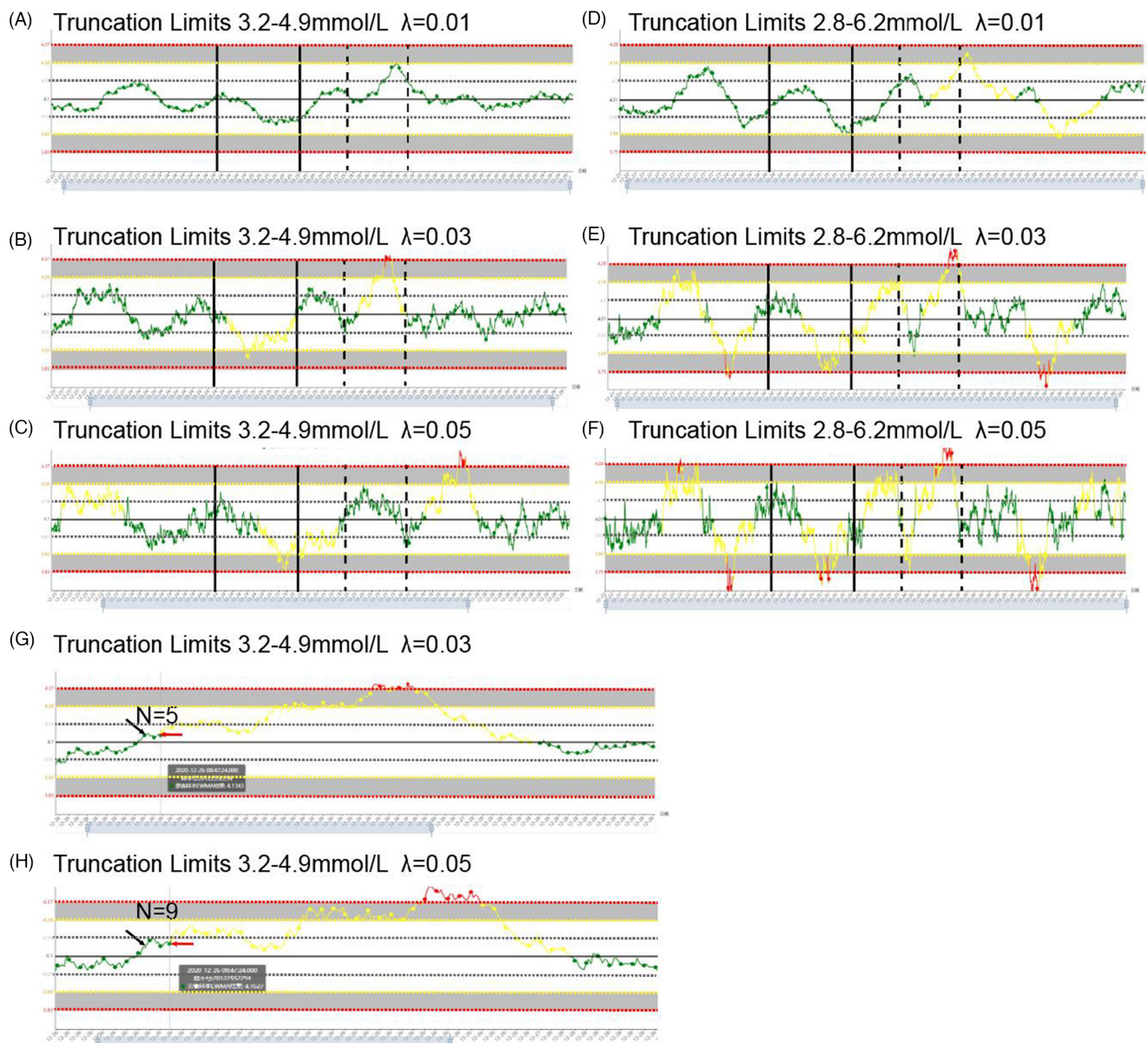


FIGURE 2 Graphic illustration of the performance of PBRTQC for potassium (K) by introducing positive (regions framed by solid lines) and negative (regions framed by dashed lines) biased data (A-F) Detected bias data at truncation limits of 3.2–4.9 mmol/L when $\lambda = 0.01$ (A) $\lambda = 0.03$ (B) and $\lambda = 0.05$ (C) and truncation limits of 2.8–6.2 mmol/L when $\lambda = 0.01$ (D) $\lambda = 0.03$ (E) and $\lambda = 0.05$ (F) Regions framed by solid and dashed lines indicate the introduced negative and positive biased data points, respectively. (G and H) Number of test results affected before error detection at truncation limits of 3.2–4.9 mmol/L when $\lambda = 0.03$ (G) and $\lambda = 0.05$ (H) The black arrows indicate the error introduction data, and red arrows indicate the error detection data. The horizontal black solid line, gray, yellow, and red dashed lines in the graph represent the mean $\pm 1, 2,$ and 3 SD of the EWMA results accumulated in the previous 6 months, respectively. The yellow and red curves of the points indicate the alarm and out-of-control data points detected, respectively, which are based on the intellectual QC rules in the AI-MA platform. Abbreviations: EMWA, exponentially weighted moving average; PBRTQC, patient-based real-time quality control; QC, quality control; SD, standard deviation

With the continuous improvement of laboratory information systems and the development of “big data” technology, professional software for the PBRTQC models was exploited and improved, promoting the application of PBRTQCs in the quality management of clinical biochemistry.^{8,18} When setting up the PBRTQC protocols, it is essential to screen out parameters through data training and to select appropriate indicators to evaluate the performance of the model, which typically requires researchers with an academic

background in computer science. Researchers have proposed algorithm optimization methods, including simulated annealing¹⁹ and grid searches,^{18,20} to enable laboratories to practically design PBRTQC procedures. In the present study, we used commercial AI-MA software to establish PBRTQC models by incorporating multiple parameters through a deep learning strategy, which was developed based on big data processing by artificial intelligence technology. Optimal PBRTQC protocols were further selected and evaluated. By

TABLE 1 Performance of PBRTQCs for serum K with different limits and smoothing factors

Truncation Limits (mmol/L)	Smoothing factor	Introduced biased data	True detected biased data	False detected biased data	Ped (%)	FPR (%)	NPed
2.8–6.2	0.01	550	176	262	33.85	13.70	— ^a
	0.03	550	476	1032	91.54	53.95	—
	0.05	550	476	1194	95.20	62.42	—
	0.10	550	56	115	11.20	6.01	—
	0.20	550	43	93	8.60	4.86	—
	0.40	550	47	105	9.40	5.49	—
3.2–4.9	0.01	550	0	0	0.00	0.00	18
	0.03^b	550	476	0	91.54	0.00	5
	0.05	550	476	203	91.54	10.61	9
	0.10	550	245	559	47.12	29.22	13
	0.20	550	40	32	7.69	1.67	16
	0.40	550	39	3	7.50	0.16	27

Abbreviations: FPR, false-positive rate; K, potassium; NPed, number of patient results affected before error detection; Ped, probability of error detection; PBRTQC, patient-based real-time quality control.

^aNot applicable.

^bOptimal PBRTQC factors are marked in bold.

using the optimal PBRTQC protocols, the comparative results were concordant with those obtained with serum samples. In general, the application of the intelligent platform may aid inspectors in more effectively selecting the optimal model for the laboratory.

Even though truncation limits were applied, patient data used for real-time monitoring systems were often affected by multiple factors, such as fluctuations in sample numbers, different sources of specimens, heterogeneous patient populations, and specific clinical interventions.²¹ Caution is particularly needed in laboratories that employ multiple instruments to measure the same analyte. The study by Song et al. implied that classifying patient data according to sources of specimens and setting quality control rules in consideration of different groups could efficiently reduce the false-positive rate of PBRTQCs.²² As influences within a group were relatively limited and did not interfere with other groups, we divided the truncated data into the inpatient (IP), outpatient (OP) and total patient (TP) groups. The negative bias detected for analyte UA when using analyzer C for the TP group was probably false, as comparative results from both serum samples and categorizing groups were excellent (Table 2). Samples run using analyzer C consisted of a major composition of inpatient samples (Figure S1). Lower results for analyte UA were observed for these samples than for those from the outpatient groups (Table S4). Thus, we speculate that the false-negative bias was likely related to the inconsistent sample numbers and deviated patient data in different groups. In addition to analyte UA, accumulated patient data for analytes ALP, GGT, K, and CA differed between the OP and IP groups. Except for analytes ALP and GGT, which were detected as having true bias, the effects from group variations for analytes K and CA might have leveled off due to the large sample quality; thus, the data in the TP group were not influenced by the aforementioned factors. In addition, in the real-time monitoring of instrument comparability, we

found that better performance in detecting small interinstrument variability and decreasing false-positive alarms was achieved using the categorized patient data. Thus, categorizing patients will be of great benefit to improve the performance of PBRTQC in comparative assays.

However, categorizing patients might lead to a decreased sample number, thus requiring further attention due to inherent limitations. For analytes with large between-subject biological variation (CV_G), the ability of PBRTQC to detect potential deviations was dramatically weakened on account of a smaller sample size and larger population variations.^{10,22} Likewise, for analytes ALT, ALP, and GGT, of which CV_G s were ranked highest, we observed false-positive and negative biased events for the OP group when using analyzer C. Nevertheless, for the analytes with relatively narrow biological variation, the desired performance was still achieved regardless of changes in sample size and patient categories. These findings emphasize the significance of considering the potential influences of low- and medium-sized populations, especially for highly varied analytes. Therefore, we suggested that PBRTQCs could be improved in relation to the sample size, population variation, and preanalytical factors for some specific analytes.

Another notable finding was that the deviated analytes were only observed when using analyzer C, which was operated 24 h a day for urgent samples. This nonstop running may have accelerated the consumption of the light source and aged the system, contributing to the instability of test samples. Enzymatic assays were particularly sensitive to manufacturing conditions, including temperature fluctuations and defective lamps,²³ which is consistent with the results that showed that analytes ALP and GGT were significantly biased when using analyzer C. In this case, the laboratory should be aware of solving this issue through more frequent calculations and earlier replacement of light sources.

TABLE 2 Results of the pairwise comparison of analyzers by serum assays and patient data

Analytes	Serum	½TE _a (%)	TP						IP						OP											
			C vs. A		C vs. B		B vs. A		C vs. A		C vs. B		B vs. A		C vs. A		C vs. B		B vs. A		C vs. A		C vs. B		B vs. A	
			EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median	EWMA	Median
ALT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
AST	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
ALP	-	† ^b	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	
GGT	-	† ^d	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	
CREA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
UA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
NA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
CA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PHOS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CA, calcium; CL, chloride; CREA, creatine; EWMA, exponentially weighted moving average; GGT, gamma-glutamyl transferase; IP, inpatient population; K, potassium; NA, sodium; OP, outpatient population; PHOS, inorganic phosphorus; TE_a, allowed total errors; TP, total population; UA, uric acid.

^aA, B, and C indicate analyzer A, B, and C, respectively.

^bIncreased data value.

^cSignificant differences are underlined in bold.

^dNo significant difference.

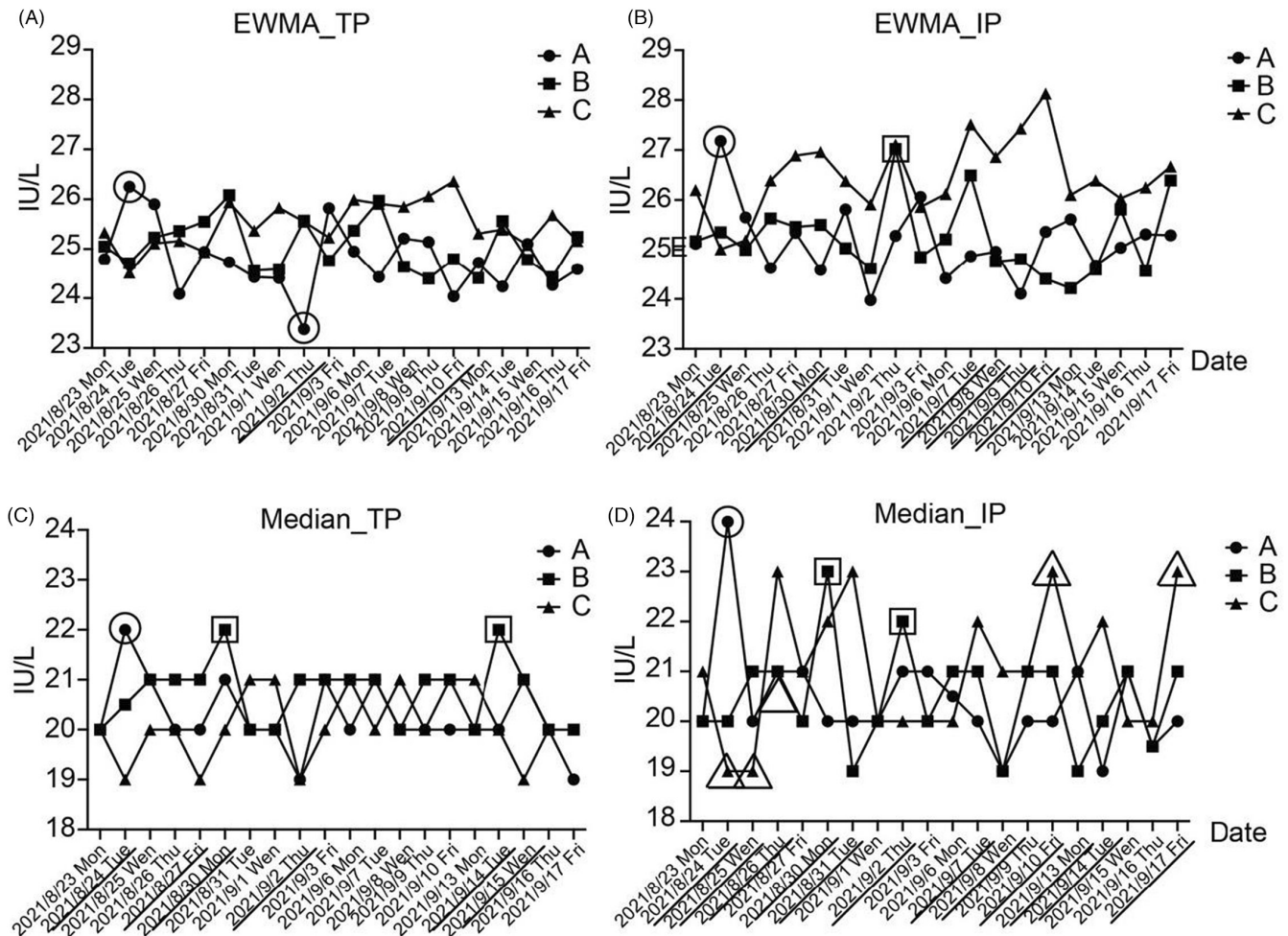


FIGURE 3 Surveillance of comparability and stability of inter- and intraanalzers for alanine aminotransferase (ALT) by (A) the EWMA method with data from the total population (TP) (B) the EWMA method with data from inpatients (IP) (C) the median method with data from TP and (D) the median method with data from IP. A, B and C indicate analyzers A, B and C, respectively. Unacceptable biases for intraanalzers were pointed out with circles for analyzer A, squares for analyzer B and triangles for analyzer C in the chart. Significant deviations for interanalzers were marked by underlining its relevant dates. Abbreviation: EWMA, exponentially weighted moving average

Since IQC and serum samples hardly reflect the stability of the detection system in the whole process, we then aimed to use the optimal EWMA and median methods to monitor the fluctuation of the inter- and intraanalzers. The EWMA algorithm appeared to perform much better than the median method in the application of daily monitoring analyzer stability, even though both methods exhibited similar performance on comparative assays. The deficiencies of the median method that led to a higher FPR became prominent when using applied patient data per day other than those from five accumulated days for comparative assays. As suggested by Kenneth Goossens on the “The Percentiler” project,¹⁰ the median method was greatly affected by population variations in small- and medium-sized samples. Therefore, PBRTQC protocols with the EWMA algorithm were more suitable as an efficient tool for the real-time monitoring of analyzer comparability and stability. However, we found that we were more likely to detect very small changes in the detection systems, for example, kit lot variations and light calibration, when the

IQCs were all in controls, and the PBRTQCs were used. We recommend that PBRTQCs, especially with an EWMA algorithm, be conducted as a supplementary procedure to expand the application of analyzer comparability to guarantee the performance of laboratory detection systems.

However, there are some limitations in our study. First, we set up individual optimal PBRTQC models with combined patient data other than the data from the analyzers. As stated by Zhou et al.²⁴ the performance of PBRTQC could be better if data were applied using separate instruments. In addition, in the simulation process, the intellectual QC rules to assess the biased data points in the chart were built in the software and were confidential for the corporation. Finally, we monitored the stability of the analyzers for a relatively short period of 1 month and only for the analyte ALT; this study should be continued for a longer time and for more analytes. In general, further studies should be conducted to optimize and expand the application of PBRTQC.

TABLE 3 Events of unacceptable bias by the EWMA and median methods with different groups

Method	Group	Date ^a	Case	Events	Comment
EWMA	TP	08/24,	A↑ ^c	A reagent lot change	True
		09/02,	A↓ ^b	None	False
		09/10	Interanalyzer	C light calibration on 9/6	True
	IP	08/24,	A↑	A reagent lot change	True
		08/30	Interanalyzer	light calibration on 8/27 8/28	True
		09/02,	B↑	B reagent lot change	True
		09/07–09/10	Interanalyzer	C light calibration on 9/6	True
Median	TP	08/24	A↑, Interanalyzer	A reagent lot change	True
		08/27	Interanalyzer	None	False
		08/30	B↑, Interanalyzer	B light calibration on 8/27	True
		09/02	Interanalyzer	B reagent lot change	True
		09/14	B↑, Interanalyzer	None	False
		09/15	Interanalyzer	None	False
	IP	08/24	A↑, Interanalyzer	A reagent lot change	True
		08/24	C↓	None	False
		08/25	C↓, Interanalyzer	None	False
		08/26	C↑, Interanalyzer	None	False
		08/30	B↑, Interanalyzer	B light calibration on 8/27	True
		08/31	C↑, Interanalyzer	C light calibration on 8/30	True
		09/02	B↑, Interanalyzer	B reagent lot change	True
		09/07–09/08	Interanalyzer	C light calibration on 9/6	True
		09/10–09/14	C↑, Interanalyzer	None	False
09/17	C↑, Interanalyzer	None	False		

Abbreviations: EWMA, exponentially weighted moving average; IP, inpatient population; TP, total population.

^aThe date of the event occurred in 2021.

^bsignificant negative biases detected.

^csignificant positive biases detected.

AUTHORS CONTRIBUTION

Yide Lu and Fan Yang designed the study and proposed the application of PBRTQC; Dongmei Wen supported the data analysis; Kaifeng Shi and Zhichao Gu collected the data and recorded cases; Qiuya Lu reviewed and revised the article; Xuefeng Wang supervised the study and reviewed the article; Danfeng Dong designed and wrote the article. All of the authors read and approved the final article.

FUNDING INFORMATION

This work was supported by the National Natural Science Foundation of China (81902117) and the Shanghai "Rising Stars of Medical Talents" Youth Development Program-Clinical Laboratory Practitioner Program.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Danfeng Dong  <https://orcid.org/0000-0002-1023-202X>

REFERENCES

- Stepman HC, Tiikkainen U, Stockl D, et al. Measurements for 8 common analytes in native sera identify inadequate standardization among 6 routine laboratory assays. *Clin Chem*. 2014;60:855-863.
- Miller WG, Ereth A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. *Clin Chem*. 2011;57:76-83.
- Hayashi S, Ichihara K, Kanakura Y, Iwatani Y. A new quality control method based on a moving average of "latent reference values" selected from patients' daily test results. *Rinsho Byori*. 2004;52:204-211.

4. Rej R, Jenny RW, Bretauidiere JP. Quality control in clinical chemistry: characterization of reference materials. *Talanta*. 1984;31:851-862.
5. Fleming JK, Katayev A. Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: for patients by patients. *Clin Biochem*. 2015;48:508-513.
6. Loh TP, Bietenbeck A, Cervinski MA, et al. Recommendation for performance verification of patient-based real-time quality control. *Clin Chem Lab Med*. 2020;58:1205-1213.
7. Bull BS, Elashoff RM, Heilbron DC, Couperus J. A study of various estimators for the derivation of quality control procedures from patient erythrocyte indices. *Am J Clin Pathol*. 1974;61:473-481.
8. Cembrowski GS, Chandler EP, Westgard JO. Assessment of "average of Normals" quality control procedures and guidelines for implementation. *Am J Clin Pathol*. 1984;81:492-499.
9. Smith FA, Kroft SH. Exponentially adjusted moving mean procedure for quality control. An optimized patient sample control procedure. *Am J Clin Pathol*. 1996;105:44-51.
10. Goossens K, Van Uytfanghe K, Twomey PJ, Thienpont LM, Participating L. Monitoring laboratory data across manufacturers and laboratories—a prerequisite to make "big data" work. *Clin Chim Acta*. 2015;445:12-18.
11. van Rossum HH, Kemperman H. Optimization and validation of moving average quality control procedures using bias detection curves and moving average validation charts. *Clin Chem Lab Med*. 2017;55:218-224.
12. Badrick T, Bietenbeck A, Katayev A, van Rossum HH, Loh TP, Cervinski MA. Implementation of patient-based real-time quality control. *Crit Rev Clin Lab Sci*. 2020;57:532-547.
13. Aarsand AK 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 <https://biologicalvariation.eu/>
14. W. Chen HS, Z. Shen, Z. Wang, C. Zhang, X. Li, F. He, X. Wang, L. Qiu, X. Liu. WS/T 403–2012 Analytical Quality Specifications for Routine Analytes in Clinical Biochemistry. 2012.
15. Institute CaLS. *Measurement Procedure Comparison and Bias Estimation Using Patient Samples*. Third ed. CLSI; 2013.
16. Danilenko U, Vesper HW, Myers GL, Clapshaw PA, Camara JE, Miller WG. An updated protocol based on CLSI document C37 for preparation of off-the-clot serum from individual units for use alone or to prepare commutable pooled serum reference materials. *Clin Chem Lab Med*. 2020;58:368-374.
17. Assessment CNASfC. Guidance on the application of accreditation criteria for the medical laboratory quality and competence in the field of Clinical Chemistry 2018.
18. Bietenbeck A, Cervinski MA, Katayev A, Loh TP, van Rossum HH, Badrick T. Understanding patient-based real-time quality control using simulation modeling. *Clin Chem*. 2020;66:1072-1083.
19. Ng D, Polito FA, Cervinski MA. Optimization of a moving averages program using a simulated annealing algorithm: the goal is to monitor the process not the patients. *Clin Chem*. 2016;62:1361-1371.
20. Duan X, Wang B, Zhu J, et al. Assessment of patient-based real-time quality control algorithm performance on different types of analytical error. *Clin Chim Acta*. 2020;511:329-335.
21. van Rossum HH. Moving average quality control: principles, practical application and future perspectives. *Clin Chem Lab Med*. 2019;57:773-782.
22. Song C, Zhou J, Xia J, Ye D, Chen Q, Li W. Optimization and validation of patient-based real-time quality control procedure using moving average and average of normals with multirules for TT3, TT4, FT3, FT3, and TSH on three analyzers. *J Clin Lab Anal*. 2020;34:e23314.
23. Cas Weykamp PF, Gunnewiek JK, de Jonge R, Kuypers A. Douwe van loon, Herman Steigstra, Christa Cobbaert, general chemistry division of the SKML (Foundation for Quality Assessment in medical laboratories, The Netherlands). Harmonisation of seven common enzyme results through EQA. *Clin Chem Lab Med*. 2014;52:1549-1555.
24. Zhou Q, Loh TP, Badrick T, Lim CY. Impact of combining data from multiple instruments on performance of patient-based real-time quality control. *Biochem Med (Zagreb)*. 2021;31(2):020705.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lu Y, Yang F, Wen D, et al. Assessment of patient based real-time quality control on comparative assays for common clinical analytes. *J Clin Lab Anal*. 2022;36:e24651. doi: [10.1002/jcla.24651](https://doi.org/10.1002/jcla.24651)