Comparison and optimization of various moving patient-based real-time quality control procedures for serum sodium

Yuanyuan Li¹,²,³ | Qian Yu¹,²,³ | Xiaoyan Zhang¹,²,³ | Xiaoling Chen¹,²,³

¹Department of Laboratory Medicine, the First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China
²Key Laboratory of Clinical In Vitro Diagnostic Techniques of Zhejiang Province, Hangzhou, China
³Institute of Laboratory Medicine, Zhejiang University, Hangzhou, China

Correspondence
Yuanyuan Li, Department of Laboratory Medicine, the First Affiliated Hospital, College of Medicine, Zhejiang University; Key Laboratory of Clinical In Vitro Diagnostic Techniques of Zhejiang Province; Institute of Laboratory Medicine, Zhejiang University, Hangzhou, China.
Email: yuanyuan2008@zju.edu.cn

Funding information
Clinical Research Fund Project of Zhejiang Medical Association, Grant/Award Number: 2016ZYC-A22

Abstract
Background: Patient-based real-time quality control (PBRTQC) is a valuable tool for monitoring the performance of testing processes. We aimed to compare and optimize various PBRTQC procedures for serum sodium.

Methods: In a computer simulation, artificial errors were added to 680,000 real patients’ results. The characteristics of error detection of various algorithms—moving average, moving median, moving SD and moving proportion of normal results including different control limits (CLs)—were assessed on their ability to detect critical errors early.

Results: The moving average and moving median were sensitive to system error, and the moving SD tended to detect random error. \( P_{3\text{SD}} \) (moving proportion of normal results, CLs based on mean and SD of proportion of normal results) demonstrated excellent performance for both system error and random error. The increase of block sizes \( N \) leads to the delay of error detection and the decrease of false rejection, except for QC procedures with minimum and maximum as CLs. CLs calculation with “0.1% false alarm rate” had more effective performance than that set false alarm to zero (minimum and maximum as CLs). The impact of truncation on QC performance depended on truncation limits, algorithms and the types of error. The significant improvement in QC performance due to truncation was only found in moving SD.

Conclusion: “\( P_{3\text{SD}} \), \( N = 50 \), without truncation” and “moving SD, \( N = 25 \), set 0.1% false alarm as CLs and set 1% outliers exclusion as truncation limits” were recommended as the optimized procedures for serum sodium to monitor system error and random error, respectively.

KEYWORDS
moving average, moving median, moving proportion of normal results, moving standard deviation, patient-based real-time quality control
1 | INTRODUCTION

Patient-based real-time quality control (PBRTQC) is a useful tool for monitoring analytic performance in clinical laboratories. It is also an important application of “big data” in laboratory quality management. Compared with traditional methods of internal quality control (QC), PBRTQC has several advantages: e.g., low cost, no matrix effect, continuous monitoring and pre-analytical monitoring. The concept of moving average QC was first published by Hoffmann and Waid in 1965. Since then, benefiting from the development of laboratory information systems, improved statistical methodology and increased awareness of the limitations of current QC, PBRTQC has attracted substantial attention and developed quickly. Recently, novel algorithms have been described, such as moving median, moving standard deviation (SD), moving average of delta and moving sum of outliers. Various parameters and charts have been developed to quantify their ability to detect error.

However, it was still a challenge for the majority of routine clinical laboratories to implement PBRTQC because of the complexity of obtaining optimal PBRTQC settings. It was well known that the probability for “error” detection of traditional individualized QC should not be less than 90%. In fact, the “error” involved here refers to “critical error”. The critical error represents the minimum error that should be detected by a QC procedure, or else it will affect clinical practice. Similarly, the optimized PBRTQC procedures should also have the best ability to detect critical error. Nevertheless, rare articles have examined the relationship between critical error and the characteristics of error detection. Furthermore, novel algorithms, such as a moving median and moving standard deviation, have been shown to be superior for error detection under certain conditions. However, most of these algorithms have been studied independently by different research groups, scarcely any articles have compared them using the same database.

In the actual application, it seemed to be more feasible that PBRTQC start with several typical tests and then be extended to all tests. Serum sodium is probably the most suitable chemistry test for PBRTQC because of its small biological variation and high requirement for analytical performance. Therefore, we aimed to investigate and compare the characteristics of error detection of various algorithms, including their different definitions of control limits for serum sodium. Both system and random error were examined, and the relationships between critical error and characteristics of error detection were described in detail to optimize PBRTQC settings.

2 | MATERIALS AND METHODS

2.1 | Patients’ data collection and errors simulation

A total of 680,000 results of serum sodium were anonymized and exported from the laboratory information system of the First Affiliated Hospital, College of Medicine, Zhejiang University, including inpatient, outpatient and physical examination population. All results were sorted by detection time and divided into 400 virtual days with 1,700 measurements each. The last 200 days served as training dataset and the first 200 days as testing dataset. All optimization of the procedures was conducted on the training dataset, and all verifications of procedure performance were conducted on the testing dataset. The robust normalized spread (RNS) was calculated on all unaltered patient measurements: RNS = interquartile range/median. RNS represents the dispersion degree of original data distribution.

Westgard JO et al. assessed the average of normal (AON) patient data algorithms to maximize run lengths for automatic process control. We used the similar method to simulate errors. The CV (analytical CV), which represents analytical inherent precision, was defined as 1/3 of the allowable total error (TEa); 1/3TEa is the minimum requirement for analytical imprecision. The TEa of sodium, which was obtained from the specification in the Analytical Quality Specification for Routine Analytics in Clinical Chemistry (WS/T 403-2012), was defined as 4%. The system error (SE) was simulated as multiples of CV (CV = 1/3TEa) by changing the mean of patients’ data. The SE ranged from 0 to 4.0 CV (0 – 4/3TEa), and both positive and negative errors were added. The random error (RE) was simulated as multiples of CV by changing the SD of the patients’ data from 1.0 to 5.0 CV. The artificial error for each day was introduced from the 201st result onwards and sustained for the remaining results.

2.2 | Parameters of QC procedures

A whole QC procedure consists of four parts: algorithms, quality control limits (CLs), truncation limits (TLs) and block size (N). Table 1 lists the QC procedures investigated in this article.

For algorithms, the moving average (A), moving median (M), moving SD (S) and the moving proportion of normal results (P) were calculated as the QC data for different error conditions and block sizes.

Several defining methods of CLs were also investigated. There were two universal methods. The one was the minimum and maximum values (mm) observed after running a calculation algorithm on the dataset without extra SE or RE. So false alarms were set to zero for these procedures. The QC procedures for this method were expressed as A mm, M mm, S mm and P mm for moving average, moving median, moving SD and the moving proportion of normal results, respectively. For the other method, the 99.95th percentile of QC data without extra error was defined as the upper CLs and 0.05th as the lower CLs. So false alarms were set to 0.1%. They were expressed as A0.1%, M0.1%, S0.1% and P0.1%.

Another two defining methods of CLs were also investigated for the moving average method (A). One was determined by calculating the reference change value (RCV) using the formula $\text{RCV} = \sqrt{2} \times Z \times \sqrt{\text{CV}_a^2 + \text{CV}_i^2}$ and CLs = mean ± RCV, where Z = 1.96 for the 2SD change in a 2-tailed distribution; CV (analytical CV)
| QC procedures | Algorithms and QC limits (CLs) | Block size (N) | Truncation (Tn) |
|---------------|-------------------------------|---------------|----------------|
| \( A_{\text{mm}, N = i, T_n} \) | Moving average (A) | CLs were the minimum and maximum values observed after running a calculation algorithm on the dataset without extra error. (mm) | Every N consecutive patients’ results were included to calculate a QC data. \( N = 25, 50, 75, 100, 125 \) and 150. |
| \( M_{\text{mm}, N = i, T_n} \) | Moving median (M) | | |
| \( S_{\text{mm}, N = i, T_n} \) | Moving SD (S) | | |
| \( P_{\text{mm}, N = i, T_n} \) | Moving proportion of normal results (P) | | |
| \( A_{0.1\%, N = i, T_n} \) | Moving average (A) | The 99.95th percentile of QC data without extra error was defined as the upper CLs and 0.05th as the lower CLs. (0.1%) | |
| \( M_{0.1\%, N = i, T_n} \) | Moving median (M) | | |
| \( S_{0.1\%, N = i, T_n} \) | Moving SD (S) | | |
| \( P_{0.1\%, N = i, T_n} \) | Moving proportion of normal results (P) | | |
| \( A_{\text{RCV}, N = i, T_n} \) | Moving average (A) | CLs = \( \text{mean} \pm \text{RCV} \) \( \text{RCV} = \sqrt{2 \times Z \times \sqrt{\text{CV}^2 + \text{CV}_i^2}} \) \( Z = 1.96 \). The \( \text{CV} \) represents analytical precision. \( \text{CV}_i \) denotes the biological variation within subjects. (RCV) | |
| \( A_{3.09, N = i, T_n} \) | Moving average (A) | CLs = \( \text{mean} \pm 3.09\text{SD}_p / \sqrt{N} \text{SD}_p \) is the SD of the patients’ data. (3.09) | |
| \( S_{C4, N = i, T_n} \) | Moving SD (S) | CLs = \( \text{SD} \pm 3 \times \text{SD}^{1-C4} \sqrt{(1-C4)\text{SD}} \) is the mean of the moving SD for an in-control period. \( C4 \) is an unbiased constant related to block size. (C4) | |
| \( P_{3SD, N = i, T_n} \) | Moving proportion of normal results (P) | CLs = \( \text{mean proportion} \pm 3 \times \text{SD proportion} \) The \( \text{mean proportion} \) is the average proportion of normal results. \( \text{SD proportion} \) is the square root of the variation for the proportion of normal results. (3SD) | |

**Note:** A whole quality control procedure (first column) consists of four parts: algorithms, quality control limits (CLs), block size and truncation limits. The italics and boldface in parentheses in the middle columns represent abbreviations.
represents analytical inherent precision; CV<sub>i</sub> (intraindividual CV) denoted the biological variation within subjects<sup>4,21</sup>. This was expressed as A<sub>RCV</sub>. The other one, which was related to block size (N), was CLs = mean ± 3.09SD<sub>p</sub>/√N<sup>19</sup>. SD<sub>p</sub> was the SD of patients’ data. This was expressed as A<sub>3DF</sub>.

The CLs for moving SD (S) were also defined with the following formula: CLs = mean ± 3 × SD<sub>proportion</sub><sup>4</sup>. Here, the mean<sub>proportion</sub> is the average proportion of normal results. The SD<sub>proportion</sub> is the square root of the variation for the proportion of normal results. It was expressed as P<sub>3SD</sub>.

To minimize the influence of outlying values in the data, truncation was usually implemented for PBRTQC protocols. According to the conclusion by Bietenbeck et al., we selected “Winsorization” method,<sup>18</sup> which replaces outlying values with the corresponding lower or upper truncation limits that was exceeded. For example, if the truncation limits were 134 – 148 mmol/L, the results greater than 148 mmol/L were replaced with 148 mmol/L instead of being eliminated directly. Three types of truncation limits (TLs), T<sub>0</sub>, T<sub>1%</sub> and T<sub>5%</sub>, were investigated. T<sub>0</sub> meant all the data were included to QC procedures, and no TLs for serum sodium. The TLs of T<sub>1%</sub> and T<sub>5%</sub> were based on the mean and SD of patients’ data (SD<sub>p</sub>). TLs of T<sub>1%</sub> was TLs = mean ± 3 × SD<sub>p</sub>, and that of T<sub>5%</sub> was TLs = mean ± 2 × SD<sub>p</sub>. T<sub>0</sub>, T<sub>1%</sub> and T<sub>5%</sub> were set to exclude the outer 0, 1 and 5% of all measurements, respectively.

We investigated QC procedures using batch sizes of 25, 50, 75, 100, 125 and 150 consecutive test results as the calculation method.

### 2.3 Performance of QC procedures

The number of patient samples necessary for error detection was counted after introducing extra error (NPed). Then median number of patient results affected before error detection (MNPed) for an increased analytical imprecision or bias was calculated. The MNPed reflected the median number of patient samples processed from the inception of an out-of-control error condition until it was detected. In addition, median number of patient samples between QC rejections when the process was in control (MNPfr) was calculated too. The MNPfr was the median number of patient samples between two false rejections. An ideal QC procedure was expected to detect error quickly and lead to rare false rejections. Thus, MNPed should be as small as possible, while MNPfr should be as large as possible. The infinite MNPeds (when the error was not detected) and MNPfrs (when false alarms were set to zero) were imputed with 1,650 (110% of the maximum value).<sup>18</sup>

### 2.4 Optimization of QC procedures

Power function graphs were generated to compare the QC procedures by plotting errors (SE and RE in the form of multiples of CV<sub>i</sub>) on the x-axis and MNPed on the y-axis.

To optimize the QC procedure, those procedures with MNPfr <1,500, which would lead to a high false rejection rate and increase QC cost, were excluded. As a false rejection is considered as a defective incident, the defective incidents per million was about 667, corresponding to a 4.75 Sigma for MNPfr ≈ 1,500.<sup>22</sup>

The ability of a QC procedure to detect critical SE and RE got particular attention. The critical system error (SEc) was calculated as follows: SEc = (TEa-bias)/CV<sub>a</sub>×1.65. In this formula, 1.65 is a z-value that sets the maximum defect rate at 5% (i.e., when the mean of patient test results has shifted by an amount that causes 5% of individual patient test results to have errors exceeding the total error requirement, the run will be considered unstable).<sup>19</sup> The critical random error (REC) was calculated as follows: REC = (TEa-bias)/1.65CV<sub>a</sub>. As all the results were exported from a stable system, bias was deemed to be zero.

An accumulative MNPed (ΣMNPed) which was the sum of MNPeds for error greater than or equal to critical error was calculated to evaluate the overall performance of QC procedure. The QC procedure (one with a minimum ΣMNPed) was selected as the optimal strategy.

All the data simulations and statistics were performed with Microsoft®Office Excel 2019 and its extended functions. The general flowchart for data simulation and QC procedure optimization is shown in Figure 1.

This study was approved by the Ethics Committee of the First Affiliated Hospital, College of Medicine, Zhejiang University.

### 3 RESULTS

#### 3.1 Data distribution

As listed in Table 2, serum sodium was nearly normally distributed with low skewness (-0.84 for training set and -0.65 for testing set) and had low RNSs (0.021 for training and 0.014 for testing set). The change of mean from training to testing set, which was the difference of means between training and testing set in relation to the mean of the training set, was -0.26%.

#### 3.2 Performance of QC procedures

A<sub>3.09</sub> and S<sub>c4</sub> were excluded because their false rejection rate was too high (MNPfr <1,500). Besides them, P<sub>3SD</sub>N = 25,T<sub>0</sub> and 1% (both MNPfr = 1137) were also ruled out for the same reason. Then the remaining rules were further assessed. The ability to detect SE and RE, which was quantified with MNPed for QC procedures under various block sizes and truncations, is shown in Table S1 and S2, respectively.
The moving average and moving median were sensitive to SE, and the moving SD tended to detect RE. Unexpectedly, $P_{3SD}$ demonstrated excellent performance for both SE and RE. Figure 2 shows the ability to detect SE, which was quantified with MNPed for QC procedures under the same block size and optimized TLs. It was demonstrated that the procedures were more capable of detecting

TABLE 2 | Statistical properties for serum sodium in the training and testing datasets

| Datasets | Mean   | SD    | Median | Max  | Min  | Skew | Kurtosis | IQR | RNS |
|----------|--------|-------|--------|------|------|------|----------|-----|-----|
| Training | 141.29 | 2.27  | 141    | 178  | 110  | −0.84| 6.36     | 3   | 0.021|
| Testing  | 140.92 | 2.25  | 141    | 185  | 108  | −0.65| 6.36     | 2   | 0.014|
| All      | 141.10 | 2.27  | 141    | 185  | 108  | −0.74| 6.24     | 3   | 0.021|

Note: Unit: mmol/L.
Abbreviations: IQR, interquartile range; RNS, robust normalized spread; SD, standard deviation.
negative SE than positive SE. This was due to a low negative skewness in original sodium distribution. P_{3SD} was clearly superior to the others in SE detection. As a whole, the A_{0.1%}, M_{0.1%} and P_{0.1%} were more competent to detect SE than A_{mm}, M_{mm} and P_{mm}, but the difference among themselves was unobvious and variable. Figure 3 shows the ability to detect RE with the same block size and optimized TLs. On the whole, the ability to detect the RE of P_{3SD}, S_{0.1%}, S_{mm}, P_{0.1%} and P_{mm} decreased in the sequence.

The influence of block size on QC performance was somewhat complex (Figure 4). The main trend was that both MNPed and MNPfr decreased with smaller block sizes. A reduction in block size led to quicker error detection, but it also led to a higher rate of false rejection. The typical cases were A_{0.1%}, A_{RCV}, P_{0.1%} and P_{3SD} for SE.
and $S_{0.1%}$ and $P_{3SD}$ for RE. As a result, the performance curves of $A_{0.1%}$ for SE and $S_{0.1%}$ for RE descended with the decrease of block size in Figure 4. However, there were exceptions for procedures with minimum and maximum as CLs. Some QC procedures with a small block size, such as $A_{mm}(N=25)$, the moving proportion of normal results with CLs based on minimum and maximum control data without extra error, $S_{0.1%}$ is the moving median number of patient samples processed from the start of an out-of-control error condition until it was detected. $CV_a$ represents analytical inherent precision. $RE$: random error. The critical $RE$ was $1.82 CV_a$. The MNPed for critical $RE$ was marked with red arrows ($\uparrow$). Figure 3 shows that the ability to detect the $RE$ of $P_{3SD}$, $S_{0.1%}$, $P_{mm}$, $P_{0.1%}$ and $P_{mm}$ decreased in the sequence.

The impact of truncation on QC performance depended on TLs, QC algorithms and the types of error (Figure 5). Truncation didn’t improve the QC performance of moving average and moving median which were sensitive to SE, but resulted in a slight increase of MNPed. Figure 2 which lists the optimized truncation limits for various procedures shows that $T_0$ was the optimal TLs for most procedures of these two algorithms. In contrast, the proper TLs can significantly
improve the QC performance of moving SD. $S_{0.1\%}$ and $S_{1\%}$ had much better performance than that without truncation (Figure 5). $T_{1\%}$ was slight superior to $T_{5\%}$. Figure 3 which lists the optimized TLs for each procedure also shows that $T_{1\%}$ was the optimal TLs for most moving SD procedures. The effect of truncation on the moving proportion of normal results depended on the TLs. If the TLs were wider than reference range, it had no impact on QC performance. If the TLs were within the reference range, its impact was fatal. For example, the TLs of $T_{1\%}$ was $134.49 - 148.07$ mmol/L, and the reference range was $137 - 147$ mmol/L. There was no difference in MNPeds between moving proportion of normal results with $T_0$ and $T_{1\%}$ (Figures 2 and 3). Conversely, the upper truncation limit of $T_{5\%}$ was $145.72$ mmol/L which was lower than upper limit of reference range ($147$ mmol/L). Compared with no truncation, the ability of $P_{mm}, P_{0.1\%}$ and $P_{3SD}$ with $T_{5\%}$ to detect RE and positive SE decreased sharply or even lost, such as $P_{3SD}, N=25$ in Figure 2A.

### 3.3 Optimized QC procedures

The critical system error (SEc) was 1.35 CVa for serum sodium. $P_{3SD}, N=50, T_{0.1\%}$ detected SEc the fastest (MNPed = 258 tests.

![Figure 4](attachment:figure4.png)

**Figure 4** Influence of block sizes on QC performance. $A_{0.1\%}$ is the moving average with 0.1% false rejection rate as CLs. $A_{mm}$ is the moving average with CLs based on minimum and maximum control data without extra error. $S_{0.1\%}$ is the moving SD with 0.1% false rejection rate as CLs. $S_{mm}$ is the moving SD with CLs based on minimum and maximum control data without extra error. MNPed: the median number of patient samples processed from the start of an out-of-control error condition until it was detected. CVa (the analytical CV) represents analytical inherent precision. $A_{0.1\%}$ and $S_{0.1\%}$ were marked with dotted lines. $A_{mm}$ and $S_{mm}$ were marked with solid lines. (A) shows the performance of $A_{0.1\%}$ and $A_{mm}$ (without truncation) for system error (SE). (B) shows the performance of $S_{0.1\%}$ and $S_{mm}$ (set 5% outliers’ exclusion as truncation limits) for random error (RE). The main trend was that both MNPed and MNPfr decreased with smaller block sizes. As a result, the performance curves of $A_{0.1\%}$ and $S_{0.1\%}$ descended with the decrease of block size in Figure 4. However, there were exceptions for procedures with minimum and maximum as CLs. Compared with $N=75$ and 125, procedures with $N=25$ detected a small error (<2.0CVa) more slowly, but detected a large error (≥2.0CVa) more quickly. That is why the performance curves of $A_{mm}, N=25$ and $S_{mm}, N=25$ intersected with that of $N=75$ and 125 in Figure 4.
for positive SEc and 33.5 tests for negative SEc). As T_0 and T_1% of P_{3SD}, N = 50 had the same performance, T_0 which was more convenient was selected. The selection of T_0, T_{1%} and T_{5%} for the other procedures followed the same way. The best 10 QC procedures for SE based on \( \Sigma MNPed \) were as follows: P_{3SD}, N = 50, T_0, \( \Sigma MNPed = 716 \); P_{3SD}, N = 75, T_0, \( \Sigma MNPed = 924 \); P_{3SD}, N = 100, T_0, \( \Sigma MNPed = 1,315.5 \); P_{3SD}, N = 125, T_0, \( \Sigma MNPed = 1,586 \); A_{0.1%}, N = 25, T_0, \( \Sigma MNPed = 2,412 \); A_{0.1%}, N = 25, T_{1%}, \( \Sigma MNPed = 2,460.5 \); M_{0.1%}, N = 50, T_0, \( \Sigma MNPed = 2,707.5 \); A_{0.1%}, N = 25, T_{5%}, \( \Sigma MNPed = 2,767 \); A_{0.1%}, N = 50, T_0, \( \Sigma MNPed = 3,861 \).

Similarly, the critical random error (REc) was 1.82CV_a. S_{0.1%}, N = 25, T_{1%} detected REc the fastest (MNPed = 24 tests). The best 10 QC procedures for RE were as follows: S_{0.1%}, N = 25, T_{1%}, \( \Sigma MNPed = 107.5 \); P_{3SD}, N = 50, T_0, \( \Sigma MNPed = 114.5 \); S_{0.1%}, N = 25, T_{1%}, \( \Sigma MNPed = 116.5 \); P_{3SD}, N = 25, T_{5%}, \( \Sigma MNPed = 132 \); P_{3SD}, N = 75, T_{0}, \( \Sigma MNPed = 152.5 \); S_{mm}, N = 25, T_{1%}, \( \Sigma MNPed = 169 \); S_{0.1%}, N = 50, T_{1%}, \( \Sigma MNPed = 171 \); S_{0.1%}, N = 50, T_{5%}, \( \Sigma MNPed = 191 \); P_{3SD}, N = 100, T_{0}, \( \Sigma MNPed = 191 \); P_{3SD}, N = 50, T_{5%}, \( \Sigma MNPed = 194.5 \); P_{3SD}, N = 125, T_{0}, \( \Sigma MNPed = 229.5 \). In all, P_{3SD}, N = 50, T_0 and S_{0.1%}, N = 25, T_{1%} were the optimized QC procedures for serum sodium, and their detailed parameters and performance are listed in Table 3.

### 3.4 The stability of the QC performance

To evaluate the stability of the PBRTQC performance over time, MNPeds from the training and the test datasets with the same
method and error were compared. Table 4 shows the SE detection performance of QC procedures in the training and test dataset. MNPeds were basically close, except for MNPeds for SE $= 1.35 CV_a$. Table 5 shows that these candidate QC procedures had highly consistent RE detection performance in the training and test dataset.

### DISCUSSION

The characteristics of the error detection of various algorithms were analyzed and compared. The moving average and moving median were sensitive to SE, and the moving SD tended to detect RE. $P_{3SD}$ demonstrated excellent performance for both SE and RE. Overall, the $A_{0.1\%}$, $M_{0.1\%}$ and $P_{0.1\%}$ were more competent to detect SE than $A_{mm}$, $M_{mm}$ and $P_{mm}$, but the difference among themselves was unobvious and variable. In general, CLs calculation with “0.1% false alarm rate” had more effective performance than that set false alarm to zero (minimum and maximum as CLs). The ability to detect the RE of $P_{3SD}$, $S_{0.1\%}$, $P_{0.1\%}$ and $P_{mm}$ decreased in the sequence. For serum sodium, $P_{3SD} = 50T_0$ and $S_{0.1\%} = 25T_1$ were the optimized QC procedures for SE and RE, respectively. $A_{3.09}$ and $S_{C4}$ were excluded for high false rejection in our research. In the previous reports, they showed acceptable, even satisfactory performance. That is because they assessed QC performance with simulated patients’ data, instead of real patients’ data. The simulated patients’ data usually have a perfect Gaussian distribution, while the real measurements probably do not follow it.

Both MNPed and MNPfr increased with the enlargement of block sizes, except for procedures with minimum and maximum as CLs. It was understandable. For example, there was a sample with $110 \text{ mmol/L}$ sodium, and it was incorporated to calculate QC data. Owing to this outlier, the average of 25 tests would obviously decrease while the average of 150 tests would just change slightly. Only after incorporating more such outliers did the QC data for $N = 150$ begin to apparently decrease. Thus, the main trend was that both MNPed and MNPfr increased with larger block sizes. Additionally, the smaller the block size, the larger the fluctuations in the QC data were. For procedures with small block sizes (such as $N = 25$), the CLs derived from minimum and maximum values were wide and not sensitive to small errors.

The impact of truncation on QC performance depended on TLs, QC algorithms and the types of error. The significant improvement in QC performance due to truncation was only found in moving SD. So only moving SD was recommended to set proper truncation limits. There are well-known significant differences between traditional QC and patient-based QC, and the limitations of traditional commercial QC have been increasingly recognized. The 4th edition of the Clinical and Laboratory Standards Institute (CLSI) C24 document

### Table 4

| Procedures | $P_{3SD}$, $N = 50$ | $P_{3SD}$, $N = 75$ | $P_{3SD}$, $N = 100$ | $P_{3SD}$, $N = 125$ | $P_{3SD}$, $N = 150$ | $A_{0.1\%}$, $N = 50$ | $M_{0.1\%}$, $N = 50$ | $A_{mm}$, $N = 50$ | $M_{mm}$, $N = 50$ |
|------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| $-4$       | 6                   | 7                   | 15                  | 27                  | 51                  | 26                  | 24                  | 31                  | 26                  |
| $-3.5$     | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $-3$       | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $-2.5$     | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $-2$       | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $-1.35$    | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $1.35$     | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $2$        | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $2.5$      | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $3$        | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $3.5$      | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |
| $4$        | 6                   | 6                   | 8                   | 27                  | 51                  | 24                  | 24                  | 30                  | 26                  |

Note: $A_{0.1\%}$ is the moving average with 0.1% false rejection rate as CLs. $A_{mm}$ is the moving average with CLs based on minimum and maximum control data without extra error. $M_{0.1\%}$ is the moving median with 0.1% false rejection rate as CLs. $M_{mm}$ is the moving median with CLs based on minimum and maximum control data without extra error. $P_{3SD}$ is the moving proportion of normal results with CLs $= \text{mean proportion} \pm 3 \times \text{SD proportion}$. All the procedures in the table had no truncation and had the same MNPfrs (1,650 tests). The MNPeds for various system errors (from 1.35CV$_a$ to 4.0CV$_a$) are listed in the table. The rows marked gray were results from training dataset, and the others were from testing set. MNPeds were basically close, except for MNPeds for SE $= 1.35 CV_a$. 

4 | DISCUSSION
recommends laboratories introduce additional QC performance metrics that are more directly related to patient risk. In other words, the traditional performance metrics (probabilities for error detection and false rejection) are not suitable for risk management. Even more important, the C24 document has proposed that the frequency of QC events and their relationship to patient risk should be the focus of QC practices. PBRTQC may be an effective way to solve these problems, due to its ability for real-time monitoring and focus directly on patients’ results. Additionally, the TEa based on desirable biological variation is usually demanding for serum calcium, chloride, sodium and albumin. The biological variation of these tests is smaller, so their TEa is stricter. As a result, the sigma metrics are so low that even multiple rules cannot achieve satisfactory performance. In contrast, the smaller the biological variation is, the more powerful the error detection ability of PBRTQC is. So, the tests with smaller biological variation are more suitable for PBRTQC, and they just need PBRTQC to make up for the inability of traditional QC to detect error. That is also why we selected sodium for our research. Nevertheless, PBRTQC is certainly complex and unpredictable compared with traditional QC. In all, these different characteristics between traditional and PBRTQC offer an opportunity to strengthen QC plans by combining them, rather than using one method in place of another.

To be specific, Figure 6 shows the proposed flowchart for the serum sodium of PBRTQC in routine clinical chemistry. Laboratory should set parameters of QC procedures firstly. The traditional individualized QC was designed based on sigma metrics of analytical performance. When setting parameters of PBRTQC, enough patient outcomes should be collected from a stable analytical system, one month at least. Then exclude outliers according to TLs, if it was needed. After that, define suitable CLs according to the optimized procedures. After proper parameters setting, the whole protocols will be performed in routine work. Traditional QC, which was usually performed at the initial phase of analysis, was applied as a confirmatory tool. If the traditional QC was in control, the analytical system started to measure patients’ samples. As measurement results were produced, PBRTQC was initiated. PBRTQC was considered as an alarm tool for monitoring performance in real time. In part, PBRTQC

| Procedures | MNPfr | 1.82 | 2   | 2.5 | 3   | 3.5 | 4   | 4.5 | 5   |
|-------------|-------|------|-----|-----|-----|-----|-----|-----|-----|
| 1650        | 22    | 19   | 16  | 14  | 12  | 12  | 11  | 11  | 11  |
| 1650        | 22    | 19   | 17  | 14  | 13  | 12  | 11  | 11  | 11  |
| 1650        | 24    | 20   | 15  | 12  | 10  | 10  | 9   | 8   | 8   |
| 1650        | 25    | 21   | 16  | 12  | 11  | 9   | 9   | 8   | 8   |
| 1650        | 27    | 23   | 16  | 12  | 10  | 10  | 9   | 8   | 8   |
| 1650        | 27    | 22   | 16  | 12  | 10  | 10  | 9   | 8   | 8   |
| 1650        | 36    | 31   | 21  | 17  | 14  | 13  | 11  | 10  | 10  |
| 1650        | 36    | 29   | 21  | 16  | 13  | 12  | 11  | 10  | 10  |
| 1650        | 38    | 33   | 24  | 19  | 16  | 15  | 14  | 13  | 13  |
| 1650        | 39    | 33   | 24  | 19  | 16  | 15  | 14  | 13  | 13  |
| 1650        | 46    | 39   | 26  | 21  | 17  | 16  | 14  | 13  | 13  |
| 1650        | 46    | 37   | 25  | 20  | 16  | 15  | 14  | 12  | 12  |
| 1650        | 56    | 47   | 32  | 25  | 20  | 18  | 16  | 16  | 16  |
| 1650        | 52    | 45   | 30  | 23  | 19  | 18  | 16  | 15  | 15  |
| 1650        | 50    | 32   | 21  | 17  | 13  | 13  | 12  | 11  | 11  |
| 1650        | 53    | 33   | 21  | 17  | 14  | 13  | 12  | 11  | 11  |
| 1650        | 55    | 44   | 32  | 25  | 22  | 19  | 18  | 17  | 17  |
| 1650        | 55    | 43   | 33  | 25  | 22  | 20  | 18  | 17  | 17  |
| 1650        | 64    | 55   | 36  | 28  | 24  | 21  | 19  | 18  | 18  |
| 1650        | 61    | 53   | 36  | 27  | 23  | 21  | 19  | 18  | 18  |
| 1650        | 140   | 76   | 29  | 23  | 20  | 19  | 18  | 17  | 17  |
| 1650        | 133   | 69   | 30  | 23  | 21  | 19  | 18  | 17  | 17  |

Note: P_{3SD} is the moving proportion of normal results with CLs = \text{mean proportion} \pm 3 \times SD_{proportion}. S_{0.1\%} is the moving SD with 0.1\% false rejection rate as CLs. S_{\text{mm}} is the moving SD with CLs based on minimum and maximum control data without extra error. T_0, T_{1\%} and T_{5\%} were set to exclude the outer 0, 1 and 5\% of all results, respectively. All the procedures in the table had the same MNPs (1650 tests). The MNPs for various random errors (from 1.82 CV_a to 5.0 CV_a) are listed in the table. The rows marked gray were results from training dataset, and the others were from testing set. These candidate QC procedures had highly consistent random error detection performance in the training and test dataset.
also decided when to perform traditional QC again. The combination of \( P_{3SD}, N = 50, T_0 \) and \( S_{0.1\%}, N = 25, T_{1\%} \) were recommended as the optimized PBRTQC for serum sodium. The detailed parameters of them are listed in Table 3. In practice, a new patient’s result corresponds to a new QC data. The large block size wouldn’t delay the startup of PBRTQC. Take \( N = 150 \) as an example, the first result of today can be combined with 149 last results of yesterday to calculate a new QC data. If PBRTQC was out of control, further measures were needed to confirm the analytical status, such as additional commercial QC or retesting retained samples.

We investigated and compared the characteristics of the error detection of various algorithms (i.e., moving average, moving median, moving SD and moving proportion of normal results), including a variety of definition methods of CLs, simultaneously. For routine laboratories, the QC procedures investigated in this paper is common and easy to implement. In addition, both SE and RE were investigated in this study. Second, the optimized QC procedure was based on the critical error instead of percentage of TEa. The critical error, which was decided by the TEa and analytical performance, was closely related to Sigma metrics of analytical system (\( SE_c = \text{Sigma metrics} - 1.65 \)). The critical error was initially applied in the designation of traditional QC procedures. Similarly, it should be of concern in PBRTQC too. Thus, it is scientifically reasonable to optimize QC procedures according to their capacity to detect
the critical error. Third, $\sum_{\text{MNPed}}$ was the advanced parameter to evaluate overall QC performance and decided the optimized QC procedures in this article. MNPed replaced ANPed as the basic parameter of QC performance. ANPed uses the average number of patient results affected before error detection, whereas MNPed uses the median number. As the numbers of results necessary for error detection is not normally distributed, MNPed is more suitable. The ability of a QC procedure to detect critical error should be received with concern. But ideally, any error greater than the critical error should be detected too. As a result, MNPed for errors greater than the critical error should be valued too. So $\sum_{\text{MNPed}}$ was more powerful than MNPed.

This study has several specific limitations. First, only serum sodium was investigated in our study, and more analytic tests with significantly different characteristics should be investigated in the future. Nevertheless, both our research and previous research$^{3,17}$ have demonstrated that serum sodium is probably the most suitable chemistry test for PBRTQC because of its small biological variation. PBRTQC is not suitable for every test, particularly, tests with low production numbers (e.g., iron), tests with an extreme variation in results (e.g., C-reactive protein and urea)$^{17}$ or a combination of both (i.e., lipase and amylase).$^{3}$ In practice, it seemed to be more feasible that PBRTQC start with several typical tests, and then be extended to most tests. Second, the series of errors were introduced using a step-shift strategy, not a gradual degradation in data simulation. In fact, a gradual degradation error may be closer to reality and be more difficult to detect.

In conclusion, the combination of $P_{1.5SD}=50,T_0$ and $S_{0.1%}N=25,T_{1.5k}$, which were the quickest to detect any type of critical error, are recommended as the optimized QC procedure for serum sodium.

ACKNOWLEDGEMENT
This work was supported by Clinical Research Fund Project of Zhejiang Medical Association (grant number 2016ZYC-A22). We thank Elixigen Corporation (Huntington Beach, California, USA) for helping in proofreading and editing the English of final manuscript.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest or financial disclosure related to this publication.

AUTHOR CONTRIBUTIONS
Yuanyuan Li took part in conceptualization, formal analysis, funding acquisition, writing the original draft and writing, reviewing and editing. Qian Yu carried out investigation and data curation. Xiaoyan Zhang was involved in methodology and data curation. Xiaoling Chen had contributed to writing, reviewing and editing.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in the supplementary material of this article.

ORCID
Yuanyuan Li https://orcid.org/0000-0001-6803-5822
Xiaoyan Zhang https://orcid.org/0000-0003-2446-8778

REFERENCES
1. van Rossum HH. Moving average quality control: principles, practical application and future perspectives. Clin Chem Lab Med. 2019;57(6):773-782.
2. Tolan NV, Parnas ML, Baudhuin LM, et al. “Big Data” in Laboratory Medicine. Clin Chem. 2015;61(12):1433-1440.
3. van Rossum HH, Kemperman H. Implementation and application of moving average as continuous analytical quality control instrument demonstrated for 24 routine chemistry assays. Clin Chem Lab Med. 2017;55(8):1142-1151.
4. Liu J, Tan CH, Badrick T, Loh TP. Moving standard deviation and moving sum of outliers as quality tools for monitoring analytical precision. Clin Biochem. 2018;52:112-116.
5. Hoffmann RG, Waid ME. The “Average of Normals” method of quality control. Am J Clin Pathol. 1965;43:134-141.
6. Hinge M, Lund ED, Brandslund I, Plesner T, Madsen JS. Patient pools and the use of “patient means” are valuable tools in quality control illustrated by a bone-specific alkaline phosphatase assay. Clin Chem Lab Med. 2016;54(3):403-406.
7. Westgard JO, Bayat H, Westgard SA. Planning risk-based SQC schedules for bracketed operation of continuous production analyzers. Clin Chem. 2018;64(2):289-296.
8. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SK. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem. 2013;59(8):1187-1194.
9. Loh TP, Cervinski MA, Katayev A, Bietenbeck A, van Rossum H, Badrick T. Recommendations for laboratory informatics specifications needed for the application of patient-based real time quality control. Clin Chim Acta. 2019;495:625-629.
10. van Rossum HH, van den Broek D. Design and implementation of average of normals (AON) patient data algorithms to incorporate the best of both worlds. Clin Chem Lab Med. 2019;57(9):1329-1338.
11. Wilson A, Roberts WL, Pavlov I, Fontenot J, Jackson B. Patient result median monitoring for clinical laboratory quality control. Clin Chim Acta. 2011;412(15–16):1441-1446.
12. Jones GR. Average of delta: a new quality control tool for clinical laboratories. Ann Clin Biochem. 2016;53( Pt 1):133-140.
13. Liu J, Tan CH, Badrick T, Loh TP. Moving sum of number of positive patient result as a quality control tool. Clin Chem Lab Med. 2017;55(11):1709-1714.
14. 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(8):1205-1213.
15. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, van Rossum HH, Loh TP. Patient-based real-time quality control: review and recommendations. Clin Chem. 2019;65(8):962-971.
16. Westgard JO. Internal quality control: planning and implementation strategies. Ann Clin Biochem. 2003;40(Pt 6):593-611.
17. 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(10):1361-1371.
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(8):1072-1083.
19. Westgard JO, Smith FA, Mountain PJ, Boss S. Design and assessment of average of normals (AON) patient data algorithms to maximize run lengths for automatic process control. Clin Chem. 1996;42(10):1683-1688.
20. Haitian Z, Chuanbao Z, Wei W, et al. Applications of two analytical specifications in the external quality assessment schemes and the external comparision of internal quality control. Chin J Lab Med. 2014;37(11):866-869.

21. Ricós C, Alvarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest. 1999;59(7):491-500.

22. Charuruks N. Sigma metrics across the total testing process. Clin Lab Med. 2017;37(1):97-117.

23. Parvin CA. 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. J Appl Lab Med. 2017;1(5):581-584.

24. Jørgensen LM, Hansen SI, Petersen PH, Soletormos G. Median of patient results as a tool for assessment of analytical stability. Clin Chim Acta. 2015;446:186-191.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

**How to cite this article:** Li Y, Yu Q, Zhang X, Chen X. Comparison and optimization of various moving patient-based real-time quality control procedures for serum sodium. J Clin Lab Anal. 2021;35:e23985. [https://doi.org/10.1002/jcla.23985](https://doi.org/10.1002/jcla.23985)