Optimization and validation of patient-based real-time quality control procedure using moving average and average of normals with multi-rules for TT3, TT4, FT3, FT3, and TSH on three analyzers

Chao Song1 | Jun Zhou2 | Jun Xia3 | Deli Ye3 | Qian Chen1 | Weixing Li1

1Zhejiang Center for Clinical Laboratories, Zhejiang Provincial People's Hospital, Hangzhou, China
2Clinical Laboratory, Hangzhou Children's Hospital, Hangzhou, China
3Clinical Laboratory Center, Zhejiang Provincial People's Hospital, Hangzhou, China

Correspondence
Weixing Li, Zhejiang Center for Clinical Laboratories, Zhejiang Provincial People's Hospital, Hangzhou, China.
Email: liweixing@hmc.edu.cn

Funding Information
This work was supported by the Soft Science Research Project of Science and Technology Department of Zhejiang Province [grant number 2019C35032]; The Medical Health Science and Technology Project of Health Commission of Zhejiang Province [grant number 2019RC016].

Abstract
Background: We have designed a patient-based real-time quality control (PBRTQC) procedure to detect analytical shifts and review analytical trends of measurement procedures.

Methods: All the nine months' patient results of total thyroxine (TT4), total triiodothyronine (TT3), free thyroxine (FT4), free triiodothyronine (FT3), and thyrotropin (TSH) measured by three identical analyzers were divided into three groups according to the source of inpatient patients, outpatient patients, and healthy people. The data in each group were truncated by optimized Box-Plot method and normalized by Box-Cox method if necessary. The z-score charts of internal quality control (IQC) samples' results and PBRTQC data were drawn by IQC levels and groups, respectively. The analytical shifts and analytical trends were detected by multi-rules of 2-2S rules and moving average rules. The performances of PBRTQC were compared with the BIQC in which IQC samples were measured only once per day at the beginning of the analytical batch. Twelve quality control cases were listed to validate the performances.

Results: All the five analytes presented normal distributions when the parameter n of Box-Plot method was 1.2. The percentages of excluded data ranged from 2.9% to 11.6%. 31 and 14 rejections triggered in PBRTQC and BIQC, respectively. 96.8% of the shift rejections in PBRTQC were trend-related shifts and calibration-related shifts, while the proportion was 85.7% in BIQC but 78.6% of the shift rejections in TSH. 25.7% and 8.6% of 105 calibration events which caused analytical shifts were detected by PBRTQC and BIQC, respectively. However, the performance of PBRTQC was not well in TSH because of its large coefficient of variation.

Abbreviations: BIQC, beginning internal quality control; CRS, calibration-related shift; CV, within-subject biologic variation; FT3, free triiodothyronine; FT4, free thyroxine; IQC, internal quality control; MA, moving averages; PBRTQC, patient-based real-time quality control; PC, positive calibration; PFR, probability for false rejection; SD, standard deviation; TRS, trend-related shift; TSH, thyrotropin; TT3, total triiodothyronine; TT4, total thyroxine; URS, unknown-related shift.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 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.

© 2020 The Authors. Journal of Clinical Laboratory Analysis published by Wiley Periodicals, Inc.

https://doi.org/10.1002/jcla.23314
1 | INTRODUCTION

The statistical quality control strategy in which stable control materials are measured in the same manner as patient specimens has been established in most laboratories to monitor their measurement procedures to detect any change relative to stable baseline. Laboratories usually define eight-hour interval per day as a stable analytical batch when the continuous quality control is performed in their automated analyzers.\(^1\) In continuous mode, the internal quality control (IQC) samples should be at least typically measured at the beginning and end of each analytical batch as well as when the reagent replenishment, calibration, or maintenance occurs. However, 50% of 1000 laboratories indicated they analyze controls only once a day at the beginning of an analytical batch.\(^2\) The percentage is much higher in China. We call this practice the beginning internal quality control (BIQC). BIQC is insufficient to rapidly detect systematic error\(^3\) which can develop at any time between IQC events and affect hundreds of patient results. The systematic error will not be detected until by the IQC event at the beginning of the next analytical batch.\(^4\) Erroneous patient results may lead to different clinical acts and thus increasing the risk of harm for patients. In addition, BIQC is designed for detection of a large, abrupt analytical shift in performance. It performs sub-optimally at detecting gradual analytical trend.\(^4,5\) Increasing IQC events frequency improves the speed of systematic error detection but also increases the cost by consumption of IQC samples, reagents, and technical time required to perform and record IQC results. Some measurement procedures are performed without IQC because of the lack of stable commercialized IQC samples. Furthermore, the IQC results are often tampered in some laboratories in China. Laboratorians prefer to select the in-control results into the IQC charts. Many IQC results beyond three times standard deviation (SD) or even two times SD are discarded or covered by repeated in-control IQC results.

The advantages of patient results-based quality control (PQC) are less matrix effect, lower cost, covering the pre-analytic quality control,\(^6,9\) etc PQC techniques include average of normals (AoN), moving averages (MA), moving median, moving sum of outliers, and annealing transformation.\(^10\) However, there are still problems for PQC on selection of normal results from patient data and eliminating the influence factors that affect the analytical performance. In this research, based on the previously published techniques, we had established a patient-based real-time quality control (PBRTQC) procedure in our laboratory to detect the analytical shifts and analytical trends at the end (or any time) of the analytical batch at a lower cost and high efficiency.

2 | MATERIALS AND METHODS

2.1 | Patient data and transformation

Analyzers and analytes: Three Beckman Coulter Dxl-800 analyzers\(^11\) (Beckman Coulter, S. Kraemer Boulevard) in Zhejiang Provincial People’s Hospital in China were selected in our research and were named ZrmDxi_1, ZrmDxi_2, and ZrmDxi_3, respectively. Five analytes including thyrotropin (TSH), total thyroxine (TT4), total triiodothyronine (TT3), free thyroxine (FT4), and free triiodothyronine (FT3) were measured in the three analyzers every day from January 1st to September 20th in 2019. An average of 400 patient specimens was analyzed every day in the three analyzers. The assembly line was used to transport specimens to the analyzers automatically. Per test, comma-separated values (CSV) export files were generated from the laboratory information system (LIS) that included the following: (a) the accurate testing time, (b) the name of the analyzer that performed the analysis, (c) the results, and (d) the source of the specimen. The files contained all available results obtained during January 1st to September 20th, 2019. The data did not involve clinical privacy of patients and other ethical issues. Two concentrations of IQC samples were measured at the beginning of the analytical batch in the morning every day in the three analyzers during the research period. The lot numbers of IQC samples were 40 341 and 40 343, which were named Lower-Level QC and Higher-Level QC, respectively. All the IQC results and testing time were exported from the laboratory information system. The data of calibration events during the research period were directly exported from each analyzer. All the units of test results were unified as TSH for mIU/L, TT3 and TT4 for nmol/L, and FT3 and FT4 for pmol/L.
Three steps of data transformations were operated as follows: First, the raw data of patient specimens during the research period were divided into the outpatient groups (OG), the inpatient groups (IG) and the healthy groups (HG) according to the patient specimens which were from the outpatient departments, the inpatient departments, or the physical examination departments. Second, the distributions of patient results in the laboratory were usually skewed because of the large number of abnormal clinical results mixed. We took the abnormal clinical results as outliers. The outliers were excluded by Box-Plot method in each group. The Box-Cox method\(^\text{12-14}\) was also used to transform the data into the normal distribution data if data were not normal distributed. Finally, all the data of a group per day were excluded if the number of included data in the group were less than 10 in that day.\(^\text{15}\) The final included data of a group per day were excluded if the number of included data were the PBRTQC data. The data transformation processes were shown. (Figure 1).

The calculation formulas of Box-Plot method were shown below (Formula 1):

\[
\text{IQR} = Q_3 - Q_1; \quad \text{Upper limit} = Q_3 + n \times \text{IQR}; \quad \text{Lower limit} = Q_1 - n \times \text{IQR}.
\]

(1)

Where \(Q_1\) was the lower quartile (the 25% quartile value, P25); \(Q_3\) was the upper quartile (the 75% quartile value, P75); IQR was the interquartile range; upper limit was the upper truncation limit; lower limit was the lower truncation limit; \(n\) was an adjustment parameter of the Box-Plot method, which was positively correlated with the range of the final included data. It was mild when \(n\) was 3 while extreme when \(n\) was 0.5. When \(n\) took different values, the numbers of weeks with normal distribution also changed. It was the optimum value of \(n\) when the maximum number of weeks with normal distribution appeared. The homogeneity of variance and normal distribution were tested by Kolmogorov-Smirnov test\(^\text{16}\) per week in each group and analyzer of the five analytes.

Daypoints: A daily unit was a set of PBRTQC data from the same analyte, group, and analyzer per day. A Daypoint was the mean of a daily units.

Moving Averages (MA): The calculation of the MA was performed as previously described.\(^\text{4,5}\) The block size of MA was set the 7 days’ Daypoints (or BIQC results).

The within-individual biological variations (CV\(_I\)) were obtained from the biological variation database of European Federation of Clinical Chemistry and Laboratory Medicine (EFLM).\(^\text{17}\) The coefficient of variations (CV\(_i\)) of Daypoints within groups and IQC results were all compared with the CV\(_I\).

### 2.2 QC rules

According to the Westgard-Sigma rules,\(^\text{18,19}\) 1-3S rules were to reject if any IQC result from the current IQC event was more than three SDs from the IQC target values and 2-2S rules were to reject if two IQC results exceed two SDs from the IQC target values in the same direction. MA rules were to reject if the MA of two groups (or two levels) exceed their control limits\(^\text{4,5}\) from the target values in the same direction at the same time. The control limit was calculated as follows (Formula 2):

\[
\text{control limit} = 3 \times \frac{\text{SD}}{\sqrt{\text{block size}}}
\]

(2)

2-2S rules were selected for Daypoints to detect the analytical shifts while MA rules were selected to detect the analytical trends. 1-3S/2-2S and MA rules were selected in BIQC. 2-2S and MA rules were selected in PBRTQC. The rules selection and possible problem suggestions were listed in the table (Table 1).

The time of calibration events did not refer to the results of BIQC or PBRTQC but was only determined by the laboratorian according to the actual status of the analyzers. When a calibration event was carried out, it was regarded as a positive calibration (PC) event if the Daypoints or IQC results changed significantly after the event or the trend of MA curves were reversed in the next few days in PBRTQC charts or BIQC charts. Otherwise, it was a negative calibration (NC) event. It was called trend-related shift (TRS) if both 2-2S rules and MA rules were triggered for several days or few days before a calibration event. It was called calibration-related shift (CRS) if 2-2S rules were triggered immediately after a PC event or the trends of MA curves reversed within few days after a PC event. The other 2-2S rules rejections were classified as unknown rejection shifts (URS).
TABLE 1 Application of multi-QC rules in PBRTQC

| Multi-rules      | Num of groups | Num of measurands | Problem suggestion         | Action                                      |
|-----------------|---------------|-------------------|----------------------------|---------------------------------------------|
| 1-3S/2-2S       | 1             | 1                 | No problem                 | No action                                  |
| 2-2S or MA      | 2             | 1                 | System deviation possible  | Keep caution                               |
| 2-2S and MA     | 2             | 1                 | System deviation confirmed | Additional IQC events or calibration verification |
| 2-2S and MA     | 2             | ≥2a               | Analyzer or pre-analytic QC | Check equipment or pre-analytic QC          |

*aThe analytes were independent of each other in the body.

2.3 | Chart drawing

2.3.1 | PBRTQC Chart

All the Daypoints were converted into z-scores. The formula (Formula 3) was shown below:

\[ z = \frac{x - \bar{x}}{SD} \]  

(3)

(x was the Daypoint; \( \bar{x} \) and SD were the mean and standard deviation of the Daypoints in the same group and analyzer during the last six months, respectively). The MA were converted into z-scores with the Formula 3. Both Daypoints and MA were plotted in a z-score chart by groups in chronological order.

2.3.2 | BIQC Chart

The test results of two IQC levels during the research period were converted into z-scores with the Formula 3. The \( \bar{x} \) and SD were the mean and standard deviation of test results during the first six months, respectively. Both test results and MA were plotted in a z-score chart by levels in chronological order.

2.4 | Cases and validation

The cases which were extracted from the rejection events in PBRTQC charts and BIQC charts were shown to compare the ability of detecting analytical deviations between PBRTQC and BIQC. It was validated that the analytical deviations were real existence if the z-scores of Daypoints or MA curves changed significantly during the handling processes of the PBRTQC rejection cases. Daily units were divided by hours. The means of the hour units were converted into z-scores with the Formula 3 and plotted in charts in the same way as the PBRTQC charts. The data in an hour unit would be excluded if the PBRTQC data were less than 5 in the hour unit.

2.5 | Statistical methods and tools

Python 3.6 was used to run the whole process. The packages of “kstest,” “Levene,” and “boxcox” in SciPy were imported for Kolmogorov-Smirnov test, homogeneity of variance test, and Box-Cox transformation. The continuous data were expressed as means ± SD. \( P < .05 \) was considered statistically significant.

3 | RESULTS

3.1 | PBRTQC data and normal distribution

The normal distribution weeks corresponding to different n values were shown in the figure (Figure 2). When n was 1.2, the largest number of normal distribution weeks appeared, accounting for 99.8% of the total 1755 weeks (5 analytes × 3 analyzers × 3 groups × 39 weeks in the research period). The percentages of

FIGURE 2 The weeks of normal distribution with the changes of n in the Box-Plot method in PBRTQC. When n was 1.2, the largest number of weeks of normal distribution appeared in the total 1755 wk (5 analytes × 3 analyzers × 3 groups × 39 wk in the research period). The maximum number of weeks was pointed by the black arrow.
excluded data in all the raw patient results ranged from 2.9% to 11.6%, with an average of 5.6%. The highest percentage of data in outpatients was discarded, with an average of 7.0 ± 2.5%. The lowest percentage of data in healthy groups was discarded, with an average of 3.7 ± 0.5%. The quality of total patient data and excluded data were listed in the table (Table 2). Data of TSH were transformed by the Box-Cox method. The PBRTQC data of the five analytes in three analyzers were normal distribution after data transformations (Figure 3).

### 3.2 QC charts and rejections

The PBRTQC charts and BIQC charts of the five analytes in three analyzers were shown in the figure (Figure 4). The rejections triggered by 1-3S and 2-2S rules were 1 and 35 times, respectively, in PBRTQC. The rejections triggered by 1-3S and 2-2S rules were 33 and 6 times respectively in BIQC. The rejections triggered by both 2-2S and MA rules were 31 and 14 times in PBRTQC and BIQC, respectively. In PBRTQC, the proportions of TRS, CRS, and URS were 83.9% (26/31), 12.9% (4/31), and 3.2% (1/31), respectively. No TRS or CRS was detected in TSH. However, the proportions of TRS, CRS, and URS were 78.6% (11/14), 7.1% (1/14), and 14.3% (2/14), respectively, in BIQC but 78.6% (11/14) of total rejections were concentrated in TSH.

In the research period, the coefficient of variations (CVs) of the five analytes in PBRTQC and BIQC were listed in the table (Table 3). In BIQC, the average of CVs of the five analytes was 6.7 ± 1.2%, and there was no significant difference between the analytes (P < .05). However, in PBRTQC, the average of CVs of TSH within the group was 22.3 ± 7.5% which was generally higher than that of other four analytes (5.0 ± 1.1%) and also significantly higher than that of TSH in BIQC (6.4 ± 1.1%). No rule rejection was triggered in TSH in PBRTQC; however, 21 times were triggered in TSH in BIQC, accounting for 53.8% (21/39) of the total rejections of BIQC.

One hundred and five calibration events occurred in three analyzers during the research period. 27(25.7%, 27/105) and 9 (8.6%, 9/105) PC events which caused analytical deviations were detected by PBRTQC and BIQC, respectively. 77.8% (7/9) of PC events detected by BIQC were also detected by PBRTQC. However, 74.1% (20/27) of PC events detected by PBRTQC could not be detected by BIQC.

### 3.3 Rejection cases and handling processes

Twelve cases about analytical deviations detected by PBRTQC in the three analyzers were listed in the table (Table 4). In PBRTQC charts, all the alarms that prompted for analytical deviations were released after subsequent processing by laboratorian and were earlier than BIQC. The twelve cases were shown to validate the ability of PBRTQC to detect the analytical deviation accumulated over a period of time or after a PC event (Figure 5).

### 4 DISCUSSION

The PQC data extracted from the results of patient specimens every day were influenced by many factors, for example the abnormal results, different sources of specimens, inconsistent number of specimens, different clinical interventions, seasonal fluctuations, specimen collections, transportations, and pre-treatments, etc. The PBRTQC established in our research also confronted these problems at least. First, the results of healthy people and patients were mixed in the laboratory; second, how to select appropriate QC rules or control limits to reduce the influence of interference factors in patient results; third, the requirement of PBRTQC on the data quantity every day.

In the previous research, specimens from diseases that had significant impacts on the data were excluded from analyzers. It was very helpful to reduce the influence of abnormal results on data, but it also increased the workload of laboratorian to screen specimens. The Committee for Standardization of Thyroid Function Tests (C-STFT) used 2 online applications, the “Percentiler” and “Flagger,” in which the medians of samples data were compared. In this research, we used the Box-Plot method to exclude the outliers from the normal distribution transformed data. It was moderate for TT3, TT4, FT3, FT4, and TSH to exclude the outliers and get the maximum number of normal distribution weeks when n = 1.2 in Box-Plot method. The percentages of excluded data in the outpatient group and the inpatient group were much higher than that of the healthy group. It might be more abnormal.
clinical results in the raw data of the outpatient group and the inpatient group. The Box-Plot method was a good way to exclude the outliers. In the process of data exclusion, it was not necessary to completely keep the included data interval consistent with the reference interval of an analyte. We tried to obtain the stable mean value or other statistical values from the large number of daily patient results to simulate IQC test results for quality control.

In our research, the PBRTQC data were divided into three independent groups according to the sources of the outpatient department, the inpatient department, and the physical examination department. The influence from different sources was limited within a group without interfering to other groups. The various interference factors within a group were regarded as independent changes within the group. The fluctuation in each day group did not only come from the analytical shifts. The rules, such as 1-3S rules, 2-2S rules, or MA rules, did not fit to quality control within a single group. The daily changes of outpatient diseases or the changes of time when inpatient specimens were sent to the laboratory might cause extreme changes in their group means. The rules triggered in only one group would increase the probability for false rejection (PFR).

In this study, 2-2S rules and MA rules were selected as the basic rules of PBRTQC to reduce the influence of interference factors between different independent groups. The 2-2S rules of Daypoints were more sensitive than MA rules to detect analytical shifts after calibration events if the significant analytical shifts occurred. 74.3% (26/35) of 2-2S rejections of Daypoints were confirmed as TRS.

**FIGURE 3** Density distribution graphs of the five analytes in three analyzers. The purple, green, and pink represented final PBRTQC data in analyzers ZrmDxi_1, ZrmDxi_2, and ZrmDxi_3, respectively. n was the parameter of Box-Plot method to exclude outliers. The data of TSH were Box-Cox transformed and presented normal distribution.

**FIGURE 4** The z-score charts of PBRTQC and BIQC. HG, healthy group; HL, higher-level IQC sample; IG, inpatient group; LL, lower-level IQC sample; MA, moving averages; NC, negative calibration; OG, outpatient group; PC, positive calibration. The z-scores of different groups in PBRTQC and IQC levels in BIQC were plotted with solid lines of different colors. The z-scores of MA of different groups in PBRTQC and IQC levels in BIQC were plotted by different black dashes. The PC events were plotted by a vertical red dotted line, while the NC events were plotted by a green vertical dotted line in both PBRTQC charts and BIQC charts. In PBRTQC charts, the z-scores triggered the multirules in two groups were marked with red dots. The z-scores only triggered the 1-3s/2-2s rules in one group were marked with black dots.
### TABLE 3  The CVs of Daypoints in PBRTQC and IQC sample results in BIQC

| Analyte | CV (%) | Analyzer | OG (%) | IG (%) | HG (%) | IQC-LL (%) | IQC-HL (%) |
|---------|--------|----------|--------|--------|--------|------------|------------|
| FT3     | 6.0    | Zrm_1    | 3.9    | 5.1    | 3.5    | 7.8        | 7.0        |
|         |        | Zrm_2    | 4.8    | 6.0    | 4.2    | 8.3        | 7.4        |
|         |        | Zrm_3    | 4.6    | 5.2    | 4.2    | 7.9        | 7.7        |
| FT4     | 7.7    | Zrm_1    | 4.4    | 4.7    | 3.7    | 6.9        | 3.5        |
|         |        | Zrm_2    | 4.3    | 4.4    | 3.3    | 6.7        | 5.1        |
|         |        | Zrm_3    | 4.2    | 4.2    | 3.4    | 7.0        | 4.0        |
| TT3     | 9.2    | Zrm_1    | 5.0    | 7.7    | 4.1    | 7.1        | 7.2        |
|         |        | Zrm_2    | 5.5    | 7.3    | 4.1    | 7.3        | 7.9        |
|         |        | Zrm_3    | 5.1    | 7.4    | 4.5    | 7.0        | 6.1        |
| TT4     | 6.4    | Zrm_1    | 4.5    | 5.7    | 4.9    | 6.4        | 5.2        |
|         |        | Zrm_2    | 6.8    | 6.9    | 6.0    | 7.0        | 6.6        |
|         |        | Zrm_3    | 5.1    | 5.3    | 4.3    | 6.6        | 4.9        |
| TSH     | 15.9   | Zrm_1    | 19.7   | 30.9   | 13.5   | 6.5        | 7.6        |
|         |        | Zrm_2    | 30.0   | 27.0   | 13.6   | 5.3        | 7.2        |
|         |        | Zrm_3    | 27.2   | 26.6   | 12.4   | 4.8        | 7.0        |

Note: CV<sup>17</sup>: the within-individual biological variation.

Abbreviations: HG, healthy group; IG, inpatient group; IQC-HL, higher-Level internal quality control sample; IQC-LL, lower-level internal quality control sample; OG, outpatient group.

### TABLE 4  Twelve cases of QC rejections in PBRTQC and handling processes

| Cases | Analyte | Analyzer | Date interval | Handling process | Figure |
|-------|---------|----------|---------------|------------------|--------|
| Case 1 | TT4     | ZrmDxi_1 | 06/01-06/30   | Calibration → PBRTQC & BIQC rejections → Recalibration → Repeated maintenance → Deviation corrected. | Figure 5A |
| Case 2 | TT4     | ZrmDxi_1 | 08/15-09/20   | System trended upward → PBRTQC warning increased → PBRTQC rejections → Recalibration → Deviation corrected. (No BIQC rejection.) | Figure 5B |
| Case 3 | TT4     | ZrmDxi_2 | 02/10-03/12   | System shifted upward → PBRTQC warning increased → PBRTQC rejections → Recalibration → Deviation corrected. (BIQC rejection lag.) | Figure 5C |
| Case 4 | TT4     | ZrmDxi_2 | 05/25-06/30   | System trended upward → PBRTQC warning increased → PBRTQC rejections → Repeated calibrations → Deviation corrected. (No BIQC rejection.) | Figure 5D |
| Case 5 | TT4     | ZrmDxi_3 | 01/01-01/31   | System trended downward → PBRTQC & BIQC rejections → Recalibration → Deviation corrected. | Figure 5E |
| Case 6 | TT3     | ZrmDxi_1 | 05/01-05/31   | Calibration → PBRTQC rejections → Recalibration → PBRTQC warning increased → Repeated maintenances → Deviation corrected. (No BIQC rejection.) | Figure 5F |
| Case 7 | TT3     | ZrmDxi_2 | 07/01-07/30   | System shifted upward → PBRTQC warning increased → recalibration → PBRTQC rejection → Repeated maintenances → Deviation corrected. (No BIQC rejection.) | Figure 5G |
| Case 8 | TT3     | ZrmDxi_3 | 04/01-04/30   | System shifted upward → PBRTQC rejection → Recalibration → Deviation corrected. (No BIQC rejection.) | Figure 5H |
| Case 9 | TT3     | ZrmDxi_3 | 05/01-06/30   | Calibration → PBRTQC warning increased → PBRTQC & BIQC rejections → Recalibration → Deviation corrected. (BIQC rejection lag.) | Figure 5I |
| Case 10 | FT3    | ZrmDxi_1 | 05/20-07/20   | System trended upward → PBRTQC warning increased → PBRTQC rejection → Recalibration → PBRTQC rejection → Repeated maintenances → Deviation corrected. (No BIQC rejection.) | Figure 5J |
| Case 11 | FT3    | ZrmDxi_2 | 01/01-02/28   | System trended downward → PBRTQC rejection → PBRTQC warning increased → Recalibration → Deviation corrected. (No BIQC rejection.) | Figure 5K |
| Case 12 | FT3    | ZrmDxi_2 | 05/01-07/20   | System trended upward → PBRTQC rejection → PBRTQC warning increased → PBRTQC rejections repeated → Recalibration → deviation corrected. (No BIQC rejection.) | Figure 5L |
MA rules could be used as a retrospective analysis and validation because it took several days to detect the analytical trend of the analyzer. It was very useful to monitor the analytical trend from the change of MA in PBRTQC and to ensure the long-term stability of analyzers. However, the block sizes and control limits of MA in different analytes were still to be adjusted. We considered that the PFR
would be reduced significantly when 2-2S and MA rules were both triggered between two independent groups in the same direction on the same day.

BIQC could monitor the stability of the detection system, but not the pre-analytic QC stages. However, PBRTQC directly reflected the quality control in the whole process from patient preparations to test results. For example, the changes of the test results caused by the failure of the cooling system of the high-speed centrifuge during the pretreatment could only be reflected in the PBRTQC. The collection, transportation, and pretreatment of specimens were all important parts of quality control in laboratories. PBRTQC was superior to BIQC in the pre-analytic QC stages. It was likely that there were problems in pre-analytic QC when PBRTQC rejections continuously if no analytical deviation was conformed in IQC.

PBRTQC was more sensitive to the deviation of test results caused by calibration events or analytical trends. 25.7% of the calibration events were prompted to cause significant changes in the analysis system by PBRTQC while the rate was 8.6% by BIQC. 96.8% of the shift rejections in PBRTQC were TRS and CRS, while the percentage of TRS and CRS was 85.7% in BIQC but 78.6% of total rejections were concentrated in TSH. It was also shown that the PC events had significant impacts on the test results through the verifications of twelve cases. The deficiencies of BIQC might be related to the irregular operation in laboratory or IQC samples. The IQC samples were only measured at the beginning of the analytical batch in BIQC and only the in-controlled test results were included artificially. For example, significant analytical deviations and several days’ rejections in TSH were not recorded at the end of July (Figure 2M-O), which might be related to the adjustment of the target value or SD of the BIQC charts. It might be the reason why the BIQC used in the laboratory was insensitive to calibration events. In contrast, PBRTQC was much less affected by laboratorians than BIQC and more realistic.

However, PBRTQC would lead to a higher PFR when the test results per day were fewer or the distribution of patients changed greatly. Therefore, we suggested that PBRTQC, as a supplementary IQC procedure, should be carried out in the laboratories with large specimens per day to reduce the cost of traditional IQC and expand the scope of IQC, especially the pre-analytic QC stages, to monitor the system deviation from the patient specimens results. It was worth noting that the trend lines of MA in PBRTQC and BIQC were the same in Figure 4C-L. PBRTQC could synchronously reflect the system deviations as BIQC at a lower cost. The ability of PBRTQC in detecting analytical deviation was significantly weakened when the CVs of PBRTQC data were large. No rejection events were triggered in TSH in PBRTQC even if the rejection events were triggered in BIQC for a long time (Figure 4 M-O). The within-individual biological variation (CV) of the analyte was the main influencing factor of the CV in PBRTQC under the same detection systems because the PBRTQC was based on the CV of the analyte. We had concerned that the detection performance of PBRTQC became weak when the CV of the analyte was significantly greater than the CV of BIQC. On the other hand, when the analyte’s CV, was not larger than the CV of its BIQC, the performance of PBRTQC was strong (Table 3). We preliminarily considered that there was a negative relationship between the CV, and the detection performance of PBRTQC. In order to control the CVS of Daypoints, we had set a minimum of 10 PBRTQC data to be included per group every day, so as to control the CVS and improve the performance of PBRTQC.

There are still some problems to be solved in PBRTQC. First, the parameters in PBRTQC need to be optimized in different analytes, for example the n in Box-Plot method, the minimum quantities of each daily group, the block size MA, and the appropriate control limit in MA. The rules will be insensitive or resulting high FPR if the parameters of MA are set too larger or too small. Therefore, further researches are needed to select the appropriate block sizes and control limits of MA. Second, PBRTQC is mainly used to detect and monitor the analytical deviations rather than random errors at present. With the development of our software, PBRTQC will also be calculated at any time in analysis process to detect analytical deviations if enough specimens were tested in two groups. Last but not least, further optimization and validation will be continued.

5 | CONCLUSION

The PBRTQC procedure can be used to monitor the analytical shifts or trends of TSH, TT3, TT4, FT3, and FT4 through the computer software and big data of patient specimens in the laboratories at a lower cost. In PBRTQC, the analytical deviations that are accumulated by systems or changed by calibration events can be triggered accurately by the multi-rules of 2-2S and MA rules. The PBRTQC can be used as a supplementary procedure to IQC every day, especially at the middle or end of the analytical batch for quality control on that day. However, the performance of PBRTQC may be weaker than that of BIQC when the within-individual biological variation of analyte is too large. The parameters in PBRTQC are still needed to adjust for each analyte.

CONFLICT OF INTEREST

The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

AUTHOR CONTRIBUTIONS

Chao Song and Jun Zhou contributed equally to this article. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

HONORARIUM

None declared.

EMPLOYMENT OR LEADERSHIP

None declared.
EXPERT TESTIMONY
None declared.

PATENTS
None declared.

ORCID
Chao Song https://orcid.org/0000-0002-5917-2224

REFERENCES
1. CLSI C24-Ed4. *Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions*, 4th ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2016.
2. Westgard SA. 2017 Global QC practices survey rest of world results. https://www.westgard.com/downloads/metrics-and-data-downloads/80-2017-global-qc-practices-survey-rest-of-world-results.html. Accessed June, 2019.
3. 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(4):473-481.
4. Liu J, Tan CH, Loh TP, et al. Detecting long-term drift in reagent lots. *Clin Chem*. 2015;61(10):1292-1298.
5. Liu J, Tan CH, Loh TP, et al. Verification of out-of-control situations detected by “average of normal” approach. *Clin Biochem*. 2016;49(16-17):1248-1253.
6. Rej R, Jenny RW, Bretaudiere JP. Quality control in clinical chemistry: characterization of reference materials. *Talanta*. 1984;31(10 Pt 2):851-862.
7. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. *Clin Chem*. 2011;57(1):76-83.
8. Ye JJ, Ingels SC, Parvin CA. Performance evaluation and planning for patient-based quality control procedures. *Am J Clin Pathol*. 2000;113(2):240-248.
9. Carobene A, Guerra E, Ceriotti F. A mechanism-based way to evaluate commutability of control materials for enzymatic measurements. The example of gamma-glutamyltransferase. *Clin Chim Acta*. 2013;424:153-158.
10. 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.
11. Wang Y, Zhang YX, Zhou YL, Xia J. Establishment of reference intervals for serum thyroid-stimulating hormone, free and total thyroxine, and free and total triiodothyronine for the Beckman Coulter DxI-800 analyzers by indirect method using data obtained from Chinese population in Zhejiang Province, China. *J Clin Lab Anal*. 2017;31(4):e22069.
12. Chaze JP. Assessing household health expenditure with Box-Cox censoring models. *Health Econ*. 2005;14:893-907.
13. Maruo K, Isogawa N, Gosho M. Inference of median difference based on the Box-Cox model in randomized clinical trials. *Stat Med*. 2015;34:1634-1644.
14. Sun L, Tong X, Zhou X. A class of Box-Cox transformation models for recurrent event data. *Lifetime Data Anal*. 2011;17:280-301.
15. Van Rossum HH, van den Broek D. Design and implementation of quality control plans that integrate moving average and internal quality control: incorporating the best of both worlds. *Clin Chem Lab Med*. 2019;57(9):1329-1338.
16. Grecu DS, Paulescu E. Quality in post-analytical phase: indirect reference intervals for erythrocyte parameters of neonates. *Clin Biochem*. 2013;46:617-621.
17. EFLM. Resources/educational materia. https://biologicalvariation.eu. Accessed September, 2019.
18. 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.
19. Schoenmakers CH, Naus AJ, Vermeer HJ, et al. Practical application of sigma metrics QC procedures in clinical chemistry. *Clin Chem Lab Med*. 2011;49(11):1837-1843.
20. Van Rossum HH. Moving average quality control: principles, practical application and future perspectives. *Clin Chem Lab Med*. 2019;57(6):773-782.
21. De Grande LAC, Goossens K, Van Uyttenhege K, et al. Monitoring the stability of the standardization status of FT4 and TSH assays by use of daily outpatient medians and flagging frequencies. *Clin Chim Acta*. 2017;467:8-14.
22. Parvin CA. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. *Clin Chem*. 2008;54(12):2049-2054.
23. Cembrowski GS, Chandler EP, Westgard JO. Assessment of “Average of Normals” quality control procedures and guidelines for implementation. *Am J Clin Path*. 1984;81(4):492-499.
24. Badrick T, Cervinski M, Loh TP. A primer on patient-based quality control techniques. *Clin Biochem*. 2019;64:1-5.

How to cite this article: 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 multi-rules for TT3, TT4, FT3, FT4, and TSH on three analyzers. *J Clin Lab Anal*. 2020;34:e23314. https://doi.org/10.1002/jcla.23314