Rui Zhou, Wei Wang, Andrea Padoan, Zhe Wang, Xiang Feng, Zewen Han, Chao Chen, Yufang Liang, Tingting Wang, Weiqun Cui, Mario Plebani* and Qingtao Wang*

# Traceable machine learning real-time quality control based on patient data

https://doi.org/10.1515/cclm-2022-0548
Received June 6, 2022; accepted June 15, 2022;
published online July 19, 2022

**Abstract**

**Objectives:** Patient-based real-time quality control (PBRTQC) has gained attention as an alternative/integrative tool for internal quality control (iQC). However, it is still doubted for its performance and its application in real clinical settings. We aim to generate a newly and easy-to-access patient-based real-time QC by machine learning (ML) traceable to standard reference data with assigned values by National Institute of Metrology of China (NIM), and to compare it with PBRTQC for clinical validity evaluation.

**Methods:** For five representative biochemistry analytes, 1,195 000 patient testing results each were collected. After data processing, independent training and test sets were divided. Machine learning internal quality control (MLiQC) was set up by Random Forest in ML and was validated by way of both metrology algorithm traceability and 4 PBRTQC methods recommended by IFCC analytical working group.

**Results:** MLiQC were established. As an example of albumin (ALB) at the critical bias, the uncertainty of MLiQC was 0.14%, which was evaluated by standard reference data produced by NIM. Compared with four optimal PBRTQC methods at critical bias, the average of the number of patient samples from a bias introduced until detected (ANPed) of MLiQC averagely decreased from 600 to 20. The median and 95 quantiles of NPeds (MNPed and 95NPed) of MLiQC were superior to all optimal PBRTQCs above 90% for all test items.

**Conclusions:** MLiQC is highly superior to PBRTQC and well-suited in real settings. The validation of the model from two aspects of algorithm traceability and clinical effectiveness confirms its satisfactory performance.

**Key words:** algorithm traceability; laboratory; patient data; real-time quality control; supervised machine learning.

Rui Zhou and Wei Wang contributed equally as co-first authors to this work.

***Corresponding authors: Mario Plebani**, Department of Medicine-DIMED, University of Padova, Padova, Italy,
E-mail: mario.plebani@unipd.it. https://orcid.org/0000-0002-0270-1711; and **Qingtao Wang**, Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China; and Beijing Center for Clinical Laboratories, No.8 Gongti South Street, Chaoyang District, Beijing, 100020, P.R. China,
E-mail: wqt36@163.com
**Rui Zhou,** Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China; and Beijing Center for Clinical Laboratories, Beijing, P.R. China
**Wei Wang,** Department of Blood Transfusion, Beijing Ditan Hospital, Capital Medical University, Beijing, P.R. China
**Andrea Padoan,** Department of Medicine-DIMED, University of Padova, Padova, Italy. https://orcid.org/0000-0003-1284-7885
**Zhe Wang, Xiang Feng and Zewen Han,** Inner Mongolia Wesure Date Technology Co., Ltd, Inner Mongolia, P.R. China
**Chao Chen,** Beijing Jinfeng Yitong Technology Co., Ltd, Beijing, P.R. China
**Yufang Liang,** Department of Laboratory Medicine, Beijing Chao-yang Hospital, Capital Medical University, Beijing, P.R. China
**Tingting Wang and Weiqun Cui,** Center for Metrology Scientific Data and Energy Metrology, National Institute of Metrology, Beijing, P.R. China

## Introduction

Patient-based real-time quality control (PBRTQC) is used as an internal quality control tool in laboratory practice [1, 2]. Through half a century, PBRTQC continues to be improved with the development of statistical methodology and information technology, and becomes an alternative/integrative tool with respect to the traditional internal quality controls (iQC) [3–5]. In 2020, a study on PBRTQC was reported by the Committee on Analytical Quality of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) [6]. However, limited by inadequate accuracy and complex algorithms, PBRTQC is still difficult to be widely adopted by medical laboratories [7, 8].

Machine learning (ML), since was put forward by Arthur Samuel in 1959, is defined as a subfield of artificial intelligence (AI) [9]. ML is good at dealing with classification or regression problems with highly accuracy, thus has inspired the development of AI-based algorithms as diagnostic or prognosis tools in healthcare areas [10–13]. As to a QC issue, our goal is to identify biased data from unbiased

data, which is belongs to ML classification issue. In the study, a novel patient-based real-time machine learning internal quality control (MLiQC) method was generated; for an example of ALB at the critical allowable bias which represented a hard-to-detect bias, MLiQC was traceable to standard reference data with assigned value by NIM to qualitatively evaluate the accuracy of the model; as a while, for examples of albumin (ALB), alanine transaminase (ALT), aspartate transaminase (AST), glucose (GLU) and total protein (TP), MLiQC method was compared with 4 PBRTQC methods recommended by IFCC to validate clinical effectiveness and utility [6]. These measurands were selected because they were the most frequently requested laboratory tests, and because they represent different distributions commonly encountered in laboratory medicine.

## Highlights

(1) A newly patient-based real-time Quality Control using machine learning (ML) is generated.
(2) MLiQC algorithm is firstly traceable to standard reference data produced by National Institute of Metrology of China (NIM), to verify the accuracy of the model.
(3) Compared with PBRTQC recommended by IFCC, especially at critical bias, the performance of MLiQC improved by above 90% for all test items.

# Materials and methods

### Patient data collection

In the period between October 2018 and July 2020, 1,195 000 patient results of five representative biochemistry test items, ALB, ALT, AST, GLU and TP were measured on Siemens Advia 2,400 were exported through Laboratory Information System (LIS) of Laboratory Department of Beijing Chao-yang Hospital, which has passed ISO15189 accreditation. Other patients-related information included age and sex were recorded. The hospital ethics board of Beijing Chao-yang Hospital approved this study (2021-D-51).

### Data simulation

As shown in Figure 1, all 1,195 000 data of each test item was kept in their original order, the first 9,56,000 data as training dataset and the rest 2,39,000 data as test dataset (the ratio of training/test datasets = 8/2) for 4 PBRTQCs/MLiQC methods. Unbiased and biased data sets were dealt with in the same way. The original data collected represented unbiased data, and then corresponding biased data was produced by introducing a bias of different sizes according to the formula (1):

$$x' = x \times (1 + P)\,(P = -50\%, -48\%, -46\%, \ldots., 46\%, 48\%, 50\%) \quad (1)$$

where $x$' represents the data after the bias introduced, $x$ is the original data, and P represents the specific relative bias value introduced in a range from –50% to 50% in the step of 2%. Therefore, 50 biased data sets of different sizes are generated for each test item based on unbiased data, covering commonly visible biases in the real laboratory settings. Due to different quality specifications for each test item, two additional critical allowable biases in positive and negative directions for each test item were introduced in order to evaluate the detection ability for hard-to-detect bias. The critical allowable bias value was calculated as the Formula (2) according to the intra-individual variation coefficient ($CV_i$) and the inter-individual variation coefficient ($CV_g$). The values of $CV_i$ and $CV_g$ for the five test items were listed (AST: 11.9%, 17.9%; ALT: 24.3%, 41.6%; GLU: 6.5%, 7.7%; TP: 2.7%, 4.0%; ALB: 3.1%, 4.2%) derived from German quality assurance plan [14].

$$\mathrm{Bias}_{\mathrm{critical}} = 0.25 \times \sqrt{CV_i^2 + CV_g^2} \quad (2)$$

In order to reproduce the process of bias detection in real laboratory setting in a short time interval, the test dataset was divided into 200 virtual days, and 1,150 data were allocated every day. The first 150 data were unbiased data, and corresponding bias of each size was introduced at 151 data site daily. Therefore, the last 1,000 data represented biased data, of which the unbiased data was set to 150 according to the maximum PBRTQC block size.

### PBRTQC

PBRTQC belonged to a statistical integrated algorithm combining multiple parameters, such as truncation limits, block sizes and algorithms. The basic principle was as follows: for each test item, 9,56,000 unbiased data included were dealt with by data truncation, data transformation and mean calculation. There were four algorithms used for mean calculation, as shown in Formulas (4)ormulas –Formulas (7). Then the control limit for each test item was calculated by using the unbiased data as a criterion for judging in-control or out-of-control status. In our study, optimal PBRTQC was used as a comparative method of MLiQC.

### Data processing for PBRTQC

Combined with the data characteristics of 9,56,000 unbiased data for each test item, firstly, a certain amount of extreme values were removed. Then the data left were dealt with by data transformation, two modes of data were obtained for each test item: without transformation and with Box-Cox transformation. Here Box-Cox transformation as shown in Formula (3), is a commonly used method of data transformation in statistics, adopting for continuous response variables which do not obey normal distribution, so as to improve the normality, symmetry and variance equality of data. It can reduce the correlation between unobservable error and prediction variables to a certain extent.

$$x'' = \begin{cases} x', & \text{neat} \\ \begin{cases} \dfrac{x'^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln x', & \lambda = 0 \end{cases}, & \text{box} - \text{cox} \end{cases} \quad (3)$$

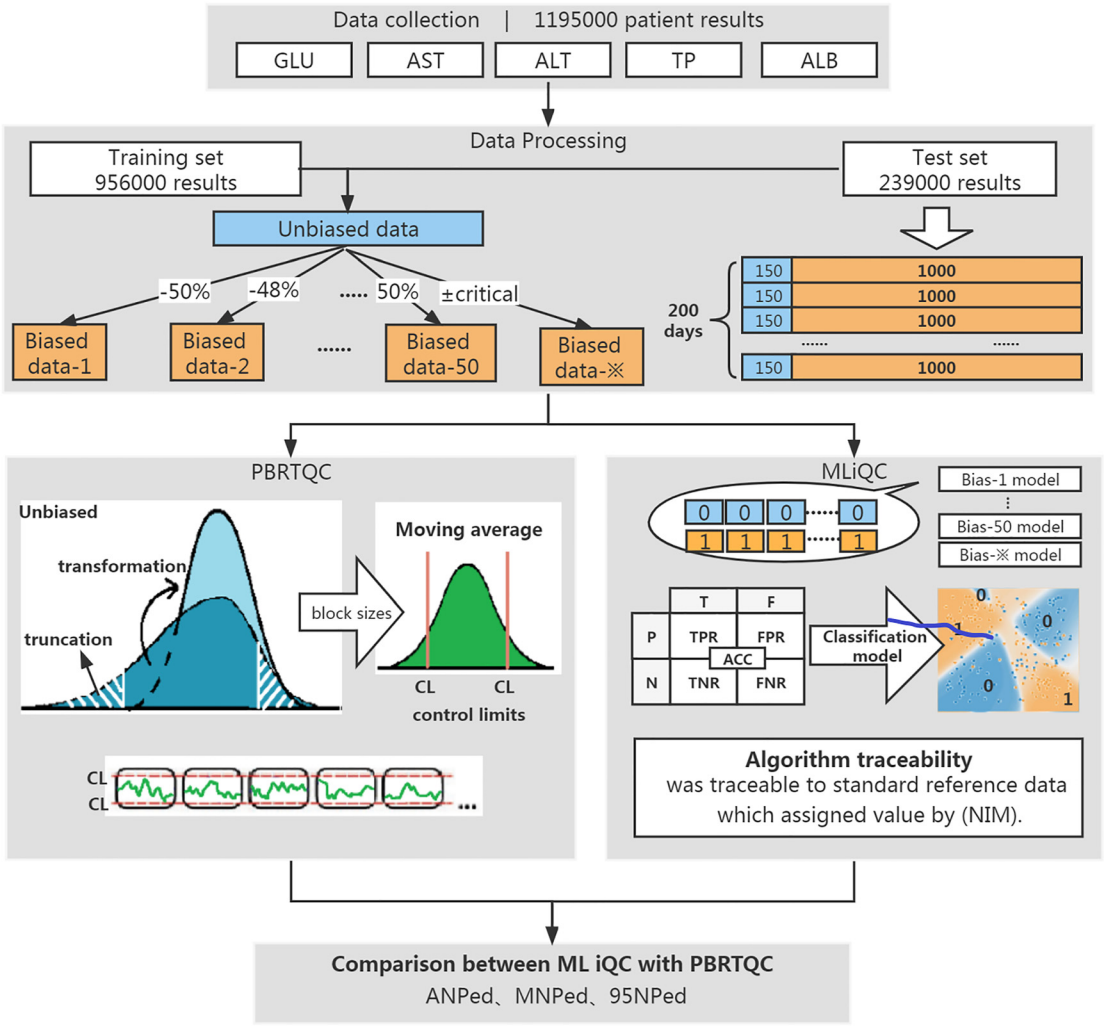


**Figure 1:** Flowchart of the experimental frame.

Where *x*' is the data before conversion, *x*'' is the data after conversion, and neat represents the data without conversion, that is, the data remained unchanged. λ is a transformation parameter, which affects the normal performance after conversion. The optimal value can be obtained directly by programming tools. The algorithms were exponentially (weighted) moving average (EMA), moving average ($MA_{mean}$), and moving median ($MM_{median}$). EMA and HD50 (15) were calculated on neat and Box-Cox transformed test results, as shown in Formula (4)ormula –Formula (7).

$$EWMA_t = \alpha \cdot MA_t + (1-\alpha) \cdot MA_{t-1} \quad (4)$$

$$MA_t = \text{mean}(x_{t-N} \sim x_t) \quad (5)$$

$$MM_t = \text{median}(x_{t-N} \sim x_t) \quad (6)$$

$$HD50_t = HD_{50}(x_{t-N} \sim x_t) = \sum_{i=1}^{N} \omega_{i:N}^{50} x_{(i)} \quad (7)$$

Where *t* refers to batch numbers of four PBRTQCs; *x* represents data amount in the current batch, and its quantity is equal to a predefined block size (n); α and ω all belong to the PBRTQC coefficient within different algorithms. These four PBRTQCs showed different performance on different clinical characteristic data, so our experiment included four PBRTQC algorithms.

## Control limits for PBRTQC

All control limits (CLs) were calculated on the training dataset of each test item and for each PBRTQC method before adding any bias. Three approaches for CLs calculation were evaluated for their tendency to produce false alarms on an intended 200 days. A relatively high but still realistic false positive alarm rate (FPR) of 5% of days was chosen. The three CLs were calculated as followed: 1) Upper and lower limits of "symmetric" CL had an equal distance to the mean of PBRTQC results, equivalent to 2SD 2) For "overall percentiles" CLs percentiles of all PBRTQC results were employed. Upper and lower limits of "overall percentiles" CLs were 0.5 and 99.5 percentiles. 3) For the "percentiles of daily extremes", the daily minimum and maximum of PBRTQC results were determined. The fifth and 95th percentiles (FPR/2) the distribution of these daily minimums and maximums defined as lower and upper CLs [6].

## MLiQC

MLiQC was aimed to separated biased data of different size from unbiased data. The steps were as followed:

Data standardization was done starting model training to guarantee data comparability. For a ML algorithm, significant data difference among testing results each test item leads to exaggerate the contribution of higher values and to weaken that of lower values. Therefore, the data of each test item is standardized according to the following formula:

$$x^* = \frac{x - \mu}{\sigma} \tag{8}$$

Here $\mu$ is the mean of all the data of a test item, $\sigma$ is the standard deviation of all the data of a test item. Because of strong randomness and indefinite characteristic attribute of the data, Random Forest (RF) as a high-performance binary classification algorithm, with strong adaptability to the data, not prone to be robust to outliers, and not overdue dependent on feature selection was employed. A detailed description of RF algorithm was given in the Supplementary Materials and Methods. Accuracy (ACC), true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), false negative rate (FNR), etc. were used for evaluating the quality of our model.

In addition, Algorithm traceability, a method for quantitative evaluation accuracy of MLiQC was introduced. MLiQC was traceable to standard reference data which was assigned value by National Institution of Metrology of China (NIM). The uncertainty of our model was evaluated according to GUM principle. An example of ALB algorithm traceability and its formula of uncertainty calculation were shown in the Supplementary Materials and Methods.

### Comparison between MLiQC and PBRTQC

The MLiQC was validated by comparing with optimal PBRTQC in 2,39,000 test dataset which were divided into 200 virtual days with 1,150 measurements daily. when FPR <5%, the number of patient samples from a bias introduced until detected was defined as NPed according to IFCC's method. Otherwise if a bias was not detected up to the end of the 1,150 Figure one day, then 1,100 value of NPed was given to programming, signaling the bias not detected [6]. In order to illustrate comprehensive clinical performance, the average, the median and 95 quantiles of NPeds (ANPed, MNPed and 95NPed) of all 200 virtual days in test dataset each test item calculated. The ways to select optimal PBRTQC methods were assessed on the training dataset of five test items. Based on the overall MNPed, the PBRTQC method with the lowest sum of MNPeds (∑MNPed) over all biases was selected. Similarly, the method with the lowest sum of 95NPeds (95NPed) was selected.

# Results

## Data distribution

For all five test items, 1,195,000 patient results were retrieved each (missingness 0%), in a time window of three years. The distribution of TP and ALB had a negative skewness (–0.82, –1.02). GLU, AST and ALT had a positive skewness (3.19, 3.21 and 2.34). The five test items examined showed skewed distribution with kurtosis from 1.07 to 7.18.

## Analytical performance of MLiQC

When false alarm rates for all five test items were less than 5% in the test dataset, the accuracy, AUC, sensitivity and specificity of MLiQC were shown in (Table 1 and Figure 2). In addition, by using reference standard data with assigned value by NIM, the uncertainty of MLiQC was only 0.14% as small as possible. It also proved the outstanding analytical performance of ML iQC from the third party. The critical bias values for the five test items were as follows: AST: 5.4%; ALT: 12%; GLU: 2.5%; TP: 1.2%; ALB: 1.3%.

## Clinical performance of MLiQC

The four optimal PBRTQCs and MLiQC for all biases and for all test items are shown in Figure 3. As an example, Figure3A shows the error detection curves for AST and ALT. In the error detection curve, if the value of the point on the curve was larger indicated that the corresponding bias detection ability was worse. The horizontal line on the top of error detection curve indicated that the bias was not detected because the bias was as small as possible, beyond the error detection ability. The wider the horizontal line on the top was, the detectable range was narrower. Two key evaluation parameters NPed and MNPed were used, respectively representing the number of patient samples from a bias introduced until detected daily and the median of NPeds of all 200 virtual days in test dataset each test item. We found that for all test items and all algorithms, the size of MNPed was obviously linear with the size of the bias introduced. It meant that with the increase of bias, the MNPed was prone to smaller. It proved the bias can be detected quicker, then the number of affecting patient results reduced.

**Table 1:** Prediction results of MLiQC at critical bias for each test item.

| Test item | FPR, % | AUC | ACC | Sensitivity | Specificity |
|---|---|---|---|---|---|
| ALT | 1.89 | 0.9889 | 0.9417 | 0.9357 | 0.9811 |
| AST | 1.03 | 0.9927 | 0.9728 | 0.9702 | 0.9897 |
| GLU | 1.29 | 0.9946 | 0.9567 | 0.9521 | 0.9871 |
| ALB | 0.45 | 0.9848 | 0.7498 | 0.7130 | 0.9955 |
| TP | 1.28 | 0.9816 | 0.7813 | 0.7504 | 0.9872 |

FPR, false positive rate; AUC, area under the receiver operating curve; ACC, accuracy; ALT, alanine transaminase; AST, aspartate transaminase; GLU, glucose (GLU); ALB, albumin; TP, total protein. The critical bias values for the five test items were as follows: AST: 5.4%; ALT: 12%; GLU: 2.5%; TP: 1.2%; ALB: 1.3%.

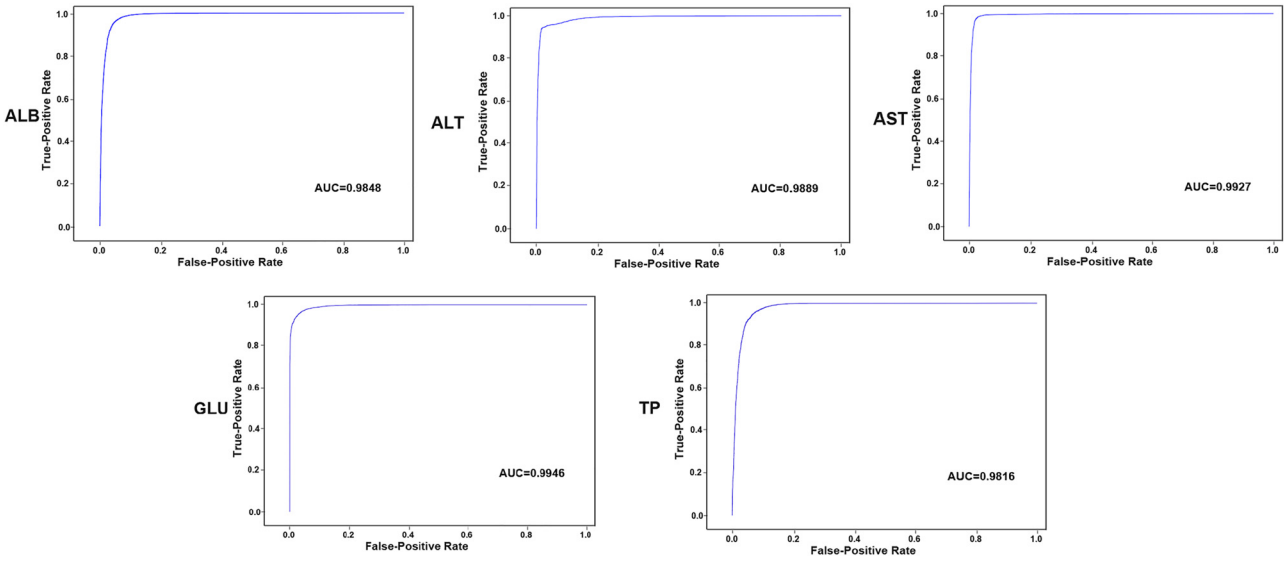


**Figure 2:** Receiver operating characteristic curves of MLiQC at critical bias for each test item.
The critical bias values for the five test items were as follows: AST: 5.4%; ALT: 12%; GLU: 2.5%; TP: 1.2%; ALB: 1.3%.

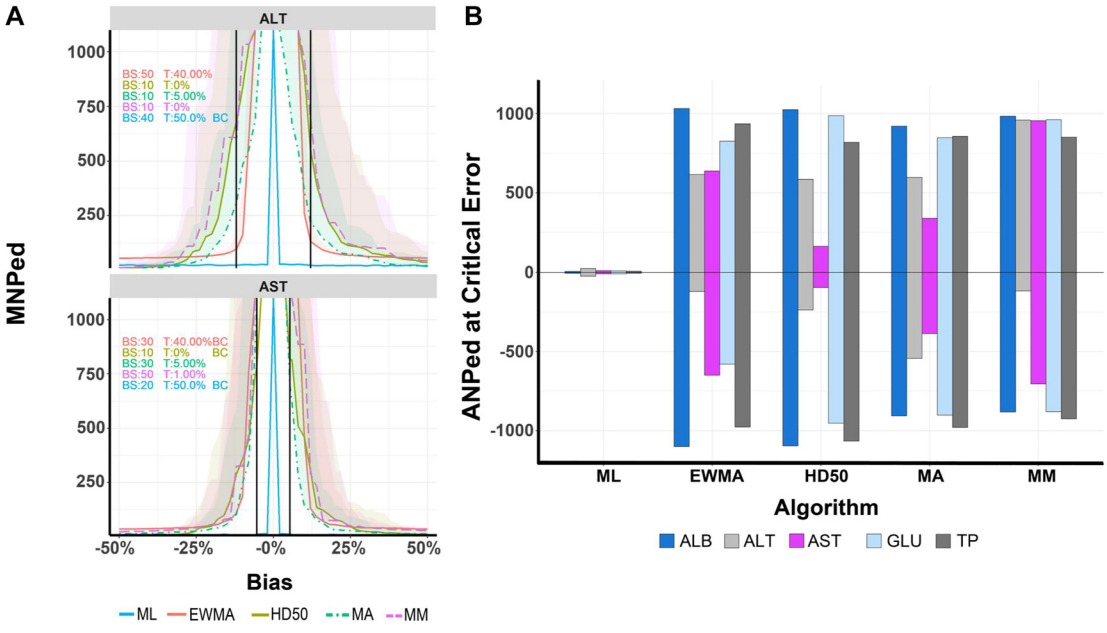


**Figure 3:** Clinical performance evaluation by comparing MLiQC with four PBRTQCs.
(A) Take ALT and AST as examples, colored lines depicted MNPed for each bias, colored area the associated 95NPed. Parameters were displayed in the top corner (BS, block size; T, truncation limit; BC, with Box–Cox transformation). Black vertical lines represented critical allowable bias each test item. (B) Histogram at the bottom represented ANPed, which represented the undetected errors for MLiQC/PBRTQCs for all test items at critical biases. Colored pillars depicted ANPed of MLiQC/ PBRTQCs for all test items at critical biases.

## Discussion

ML has gained a place in medicine and has captured the interest of medical researchers and practitioners within this subfield of computer science [15]. The enormous technological potential developed over the last years is increasingly influencing life sciences and driving changes toward personalized medicine and value-based healthcare [16, 17]. In previous studies, ML applications in laboratory medicine focused on the establishment of reference intervals, diagnostic prognostic models, epidemiological investigation, and analysis of sources of variations in analytes [18–20]. To our knowledge, seldom studies have been reported aiming to develop a quality control approach in laboratory medicine. We aimed to generate a novel quality control approach by using supervised ML for analytical error detection in real settings, with consideration to cost, appropriateness, easy access, and reproducibility of measurements. In our study, 50 biases of different sizes were stimulated and 200 visual days were divided in test sets based on statistical principles, in order to reproduce different clinical scenes within a short time interval.

Compared with IFCC's PBRTQC in Figure 3, we found that the smaller the bias was, the larger the MNPed was. MLiQC was more sensitive and stable to four optimal PBRTQCs for all biases and all test items in the parameters of MNPed, 95NPed Figure 3A and ANPed Figure 3B. In each curve, ANPed、MNPed、95NPed showed definite cutoff for all algorithms. The cutoffs of four PBRTQCs for all test items were between 20 and 25% in absolute value. Then with the shrinking of bias introduced which was lower than 20% in absolute value, MNPed of four PBRTQCs increased exponentially. However, the cutoff of MLiQC was close to 2% in absolute value. MLiQC could detect all biases more than 2% in absolute value, and its MNPed was stable, below 20 for all test items. 95NPed and had the similar change trend like MNPed in two types of methods. Asymmetrical error detection curves could be observed in ALT and AST for biases in both directions for four PBRTQCs, but MLiQC did not show significant difference for all test items. Especially for hard-to-detect critical bias each test item, MNPed of PBRTQC fell into the positive and negative cutoff range, that indicted MNPed was unable. Otherwise, the MNPed of MLiQC was low and stable in Figure 3A. ANPed of MLiQC in the histogram in Figure 3B. performed similar trend as MNPed and 95NPed, it averagely decreased from 600 to 20, improving by above 90% than four PBRTQCs for all test items. Take for an example of ALT representing extreme data disturbance, Figure 3. showed that whenever the systematic error changed, the MLiQC outperformed the four PBRTQCs.

It was also the first time in our study to put forward to metrological traceability to validate the accuracy of MLiQC. In the point of view of metrology, measurement traceability is the property of a measurement whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty [21]. MLiQC was traceable to standard reference data which is the data related to a property of a phenomenon, body or substance, or to a system of components of known composition or structure, obtained from an identified source, critically evaluated and verified for accuracy, and issued by national authority [21]. The standard reference data of MLiQC was produced and assigned value by NIM in accordance with the Metrology Law of the People's Republic of China and Standard Reference Data Act of the United States [22]. It offers higher hierarchy method for quantitative evaluation of the MLiQC. Consequently, it can ensure the accuracy and reliability of the MLiQC in real testings.

The interpretability of MLiQC should be taken into consideration. In 2018, the Department of Information Industry under the United States Food and Drug Administration (FDA) proposed at the regulatory level that the market access criteria for the development of ML products using medical data should be evaluated from three aspects [23]: algorithm program effectiveness, clinical effectiveness and clinical applicability. Of all 1,195,000 patient results in the period of October 2018 and July 2020, retrieved each test item (missingness 0%) kept in their original order, the first 9,56,000 data was as training dataset and the rest 2,39,000 data as test dataset. By comparing with clinical recognized four PBRTQCs in test dataset, clinical effectiveness of MLiQC was evaluated in comparison with IFCC's PBRTQC method. Furthermore, the results of proficiency testing organized by Beijing Center for Clinical Laboratory last year didn't show significant difference among peer groups for the five test items. It is also proved MLiQC is suitable to clinical applicability from clinical testing result compatibility. In addition, initially introduced, algorithm traceability adopts third party standard reference data generated by NIM for further validating the reliability of MLiQC. The limitation of our study, we just compared MLiQC to four PBRTQCs by artificial error data to validate our model. If available, they are likely to be implemented in laboratory information systems for further validation in the context, finally offering an easy-to-access software and helping to the improve the quality of laboratory test results.

## Conclusions

MLiQC is highly accurate and fast for analytical error detection of various sizes occurred in real laboratory settings. We provide a new thinking for ML applications for unreliable patient results prediction.

**Research funding:** This work was supported by two funding programs: the National Clinical Key Specialty Construction Projects and High-tech industry major innovation support projects (CYGX2112) for "Development a new quality control platform based on artificial intelligence using routine laboratory test data".

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Informed consent:** Not applicable.

**Ethical approval:** Not applicable.

## References

1. Hoffmann RG, Waid ME. The "AVERAGE OF NORMALS" method of quality control. Am J Clin Pathol 1965;43:134–41.
2. 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;61:473–81.
3. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, Van Rossum HH, Loh TP. International federation of clinical chemistry, and laboratory medicine committee on analytical quality. Patient-based real-time quality control: review and recommendations. Clin Chem 2019;65:962–71.
4. Westgard JO, Bayat H, Westgard SA. Planning risk-based SQC schedules for bracketed operation of continuous production analyzers. Clin Chem 2018;64:289–96.
5. Rossum HHV, 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;55:1142–51.
6. 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;66:1072–83.
7. Cembrowski GS, Chandler EP, Westgard JO. Assessment of "Average of Normals" quality control procedures and guidelines for implementation. Am J Clin Pathol 1984;81:492–9.
8. van Rossum HH. Moving average quality control: principles, practical application and future perspectives. Clin Chem Lab Med 2019;57:773–82.
9. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev 1959;3:210–29.
10. Bennie M, Malcolm W, McTaggart S, Mueller T. Improving prescribing through big data approaches-Ten years of the Scottish Prescribing Information System. Br J Clin Pharmacol 2020;86:250–7.
11. Ma C, Wang X, Wu J, Cheng X, Xia L, Xue F, et al. Real-world big-data studies in laboratory medicine: current status, application, and future considerations. Clin Biochem 2020;84:21–30.
12. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319:1317–8.
13. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, et al. Deep learning and process understanding for data-driven Earth system science. Nature 2019;566:195–204.
14. German Medical Association. Revision of the "guideline of the German medical association on quality assurance in medical laboratory examinations–rilibaek". J Lab Med 2015;39:26–69.
15. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med 2018;56:516–24.
16. 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; 62:1361–71.
17. van Rossum HH, Kemperman H. Moving average for continuous quality control: time to move to implementation in daily practice? [Letter]. Clin Chem 2017;63:1041–3.
18. Duan X, Wang B, Zhu J, Zhang C, Jiang W, Zhou J, et al. Regression-adjusted real-time quality control. Clin Chem 2021; 67:1342–50.
19. U.S. Food and Drug Administration. Use of real-world evidence to support regulatory decisions-making for medical devices: guidance for industry and Food and Drug Administration staff; 2017. Available from: https://www.fda.gov/media/99447/download.
20. Price W, Cohen IJNM. Privacy in the age of medical big data. Nat Med 2019;25:37–43.
21. International vocabulary of metrology. Basic and general concepts and associated terms (VIM), ISO/IEC GUIDE 99. Sèvres: Joint Committee for Guides in Metrology; 2012.
22. Standard Reference Data Act. Public law 90-396. United States Congress; 1968. Available from: https://www.nist.gov/system/files/documents/srd/SRDAct-2.pdf.
23. Khoury MJ, Ioannidis JP. Medicine. big data meets public health. Science 2014;346:1054–5.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2022-0548).