Clinical Biochemistry xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

# Moving standard deviation and moving sum of outliers as quality tools for monitoring analytical precision

Jiakai Liu[a], Chin Hon Tan[a], Tony Badrick[b], Tze Ping Loh[c,d,*]

[a] Department of Industrial and Systems Engineering, National University of Singapore, Singapore
[b] Faculty of Health Sciences and Medicine, Bond University, Queensland, Australia
[c] Department of Laboratory Medicine, National University Hospital, Singapore
[d] Biomedical Institute for Global Health Research and Technology, National University of Singapore, Singapore

ARTICLE INFO

*Keywords:*
Quality control
Moving standard deviation
Moving average
Average of normal
Moving sum
Analytical error
Imprecision
Erroneous
Spurious
Laboratory error
Average number of patient results affected before error detection
Coefficient of variation
Analytical coefficient of variation
Moving sum of outlier
Laboratory Management
Abbreviations
AoN
average of normal
ANPed
average number of patient results affected before error detection
CV
coefficient of variation
CVa
analytical coefficient of variation
moving SD
moving standard deviation
movSO
moving sum of outlier
QC
quality control
SD
standard deviation

ABSTRACT

*Introduction:* An increase in analytical imprecision (expressed as $CV_a$) can introduce additional variability (i.e. noise) to the patient results, which poses a challenge to the optimal management of patients. Relatively little work has been done to address the need for continuous monitoring of analytical imprecision.

*Methods:* Through numerical simulations, we describe the use of moving standard deviation (movSD) and a recently described moving sum of outlier (movSO) patient results as means for detecting increased analytical imprecision, and compare their performances against internal quality control (QC) and the average of normal (AoN) approaches.

*Results:* The power of detecting an increase in $CV_a$ is suboptimal under routine internal QC procedures. The AoN technique almost always had the highest average number of patient results affected before error detection (ANPed), indicating that it had generally the worst capability for detecting an increased $CV_a$. On the other hand, the movSD and movSO approaches were able to detect an increased $CV_a$ at significantly lower ANPed, particularly for measurands that displayed a relatively small ratio of biological variation to $CV_a$.

*Conclusion:* The movSD and movSO approaches are effective in detecting an increase in $CV_a$ for high-risk measurands with small biological variation. Their performance is relatively poor when the biological variation is large. However, the clinical risks of an increase in analytical imprecision is attenuated for these measurands as an increased analytical imprecision will only add marginally to the total variation and less likely to impact on the clinical care.

## 1. Introduction

Laboratory results are most often performed to monitor the progression of disease in a patient. An increase in analytical imprecision can introduce additional variability (i.e. noise) to the patient results [1]. This can impair the accurate assessment of the underlying trend in

* Corresponding author at: Department of Industrial and Systems Engineering, National University of Singapore, Singapore.
*E-mail address:* tploh@hotmail.com (T.P. Loh).

http://dx.doi.org/10.1016/j.clinbiochem.2017.10.009
Received 8 September 2017; Received in revised form 13 October 2017; Accepted 15 October 2017

the patient results. It can also increase the probability of erroneous classification of patients when a clinical decision threshold is applied. Together, they pose a challenge to the optimal management of patients.

The routine performance of laboratory tests is commonly monitored by measurement of internal quality control (QC) samples at fixed intervals or at the beginning and/or the end of an analytical run. However, the internal QC system is designed for detection of a large, abrupt analytical shift in performance. It performs suboptimally at detecting gradual analytical drift [2,3]. Yago has recently described statistical procedures for selecting appropriate internal QC rules for detection of increased analytical imprecision [4]. While they are reasonably simple to use and has good performance (provided sufficient frequency/number of internal QC samples are measured), they still rely on samples which may not be directly commutable with clinical samples.

There is increasing call for the use of alternate risk-based statistical tools, where their performance is directly linked to patients who are at risk/being affected by the analytical error, to better monitor the clinical impact and analytical performance of instrument in the high throughput setting of laboratory medicine practice today [4–6]. In particular, the use of successive patient results as a continuous feedback on instrument performance is gaining acceptance. One example of such statistical technique is the moving average of patient results, which has been shown to detect analytical shift well [7–9]. By contrast, relatively little work has been done to address the need for continuous monitoring of analytical imprecision. In this study, we describe the use of moving standard deviation and a recently described moving sum of outlier patient results [10] as means for detecting increased analytical imprecision, and compare their performances against internal QC and the average of normal (AoN) approaches.

## 2. Methods

### 2.1. Ability of internal quality control system to detect an increased imprecision

Firstly, the ability of routine internal QC procedures to detect an increase in analytical imprecision (expressed as analytical coefficient of variation, $CV_a$) under various internal (Westgard-based) QC rules is examined. We assume the internal QC results followed a Gaussian distribution and there is an increase in $CV_a$ for the internal QC samples reflecting an increase in random error. The corresponding standard z-values for the probability of detecting the increased $CV_a$ for the routine QC procedures under 1:1S, 1:2S and 1:3S rules are shown as follows:

$$z_{1:1S} = \frac{1 \times \mu \times CV_{old}}{\mu \times CV_{new}} = \frac{CV_{old}}{CV_{new}}$$

$$z_{1:2S} = \frac{2 \times CV_{old}}{CV_{new}}$$

$$z_{1:3S} = \frac{3 \times CV_{old}}{CV_{new}}$$

where $\mu$ denotes the target value for the internal QC sample, while $CV_{old}$ and $CV_{new}$ denote the original $CV_a$ and the increased $CV_a$, respectively. The corresponding probability of detecting an increased $CV_a$ under a specific QC rule can be obtain by referring to the z-table (see worked example below).

Let $P_1$,$P_2$ and $P_3$ denote the probability of having a single measurement exceeding the 1SD, 2SD and 3SD limits, respectively. The probabilities of detecting an increase in $CV_a$ for the internal QC procedures under 1:2S, 2:2S, 1:3S, 4:1S rules, denoted as $P_{1:2S}$, $P_{2:2S}$, $P_{1:3S}$, $P_{4:1S}$, respectively, are:

$$P_{1:2S} = P_2 \times 2$$

$$P_{2:2S} = (P_2)^2 \times 2$$

$$P_{1:3S} = P_3 \times 2$$

$$P_{4:1S} = (P_1)^4 \times 2$$

For example, a 50% increase in $CV_a$ will trigger the 2:2S rule with a probability of 1.7%, or once every 60 consecutive internal QC runs. This can be obtained by first calculating the standard z-value for the 1:2S rule as:

$$z = \frac{2 \times cv}{cv \bullet (1 + 0.5)} = 1.33$$

The probability of having a result smaller than 2SD limit can be obtained by referring to the z-table, which is 0.9082 in this case. Thus, the probability of having one result larger than 2SD at this magnitude of increase in $CV_a$ is$1 - 0.9082 = 0.0918$. The probability ofobtaining two consecutive internal QC results exceeding 2SD (2:2S rule) is $2 \times 0.0918^2 = 0.1667$ ($or \sim 1.7\%$).

Additionally, we examined the power of error detection when two or three QC samples at different concentrations are run at the same time. The rejection is made if one or more samples exceed the predefined control limit (2SD or 3SD). Here, the probabilities of QC rejection are denoted as $P_{2/1:2S}$, $P_{2/1:3S}$, $P_{3/1:2S}$ and $P_{3/1:3S}$. They are calculated as:

$$P_{2/1:2S} = 1 - (1 - P_{1:2S})^2$$

$$P_{2/1:3S} = 1 - (1 - P_{1:3S})^2$$

$$P_{3/1:2S} = 1 - (1 - P_{1:2S})^3$$

$$P_{3/1:3S} = 1 - (1 - P_{1:3S})^3$$

The $R_{4S}$ ($R_{2S}$) refers to the control rule when two QC samples are measured at the same time, the QC run is rejected when one QC measurement exceeds the mean + 2SD (+1SD) while the other exceeds the mean − 2SD (−1SD), or vice versa. The probability of triggering the $R_{4S}$ ($R_{2S}$) rule (denoted as $P_{R_{4s}}$ and $P_{R_{2s}}$) is:

$$P_{R4s} = P_2^2 \times 2$$

$$P_{R_{2s}} = P_1^2 \times 2$$

The probabilities of violating different routine internal QC rules in the presence of an increase in $CV_a$ by 50%, 100% and 200% are summarized in Table 1.

**Table 1**
Probabilities of triggering different internal quality control (QC) rules and the estimated number of QC runs it takes to detect a 50%, 100%, 200% increase in analytical imprecision.

| QC rule | Power of detection | | | No. of QC runs to detection | | |
|---|---|---|---|---|---|---|
| | $CV_a \uparrow 50\%$ | $CV_a \uparrow$ 100% | $CV_a \uparrow$ 200% | $CV_a \uparrow$ 50% | $CV_a \uparrow$ 100% | $CV_a \uparrow$ 200% |
| One QC sample/run | | | | | | |
| 1:2S | 18.3% | 31.7% | 50.5% | 5 | 3 | 2 |
| 1:3S | 4.6% | 13.4% | 31.7% | 22 | 7 | 3 |
| 4:1S | 0.8% | 1.8% | 3.7% | 123 | 55 | 27 |
| Two QC samples/run | | | | | | |
| 1:2S | 33.2% | 53.4% | 75.5% | 3 | 2 | 1 |
| 2:2S | 1.7% | 5.0% | 0.8% | 60 | 20 | 126 |
| 1:3S | 8.9% | 24.9% | 53.4% | 11 | 4 | 2 |
| 4:1S | 1.6% | 3.6% | 7.3% | 62 | 28 | 14 |
| R4S | 1.7% | 5.0% | 12.8% | 60 | 20 | 8 |
| R2S | 12.8% | 19.0% | 27.3% | 8 | 5 | 4 |
| Three QC samples/run | | | | | | |
| 1:2S | 45.4% | 68.2% | 87.9% | 2 | 1 | 1 |
| 1:3S | 13.0% | 35.0% | 68.2% | 8 | 3 | 1 |
| 4:1S | 2.4% | 5.3% | 10.8% | 41 | 19 | 9 |