Review

Huub H. van Rossum*

Moving average quality control: principles, practical application and future perspectives

https://doi.org/10.1515/cclm-2018-0795
Received July 25, 2018; accepted September 10, 2018; previously published online October 11, 2018

Abstract: Moving average quality control (MA QC) was described decades ago as an analytical quality control instrument. Although a potentially valuable tool, it is struggling to meet expectations due to its complexity and need for evidence-based guidance. For this review, relevant literature and the world wide web were examined in order to (i) explain the basic concepts and current understanding of MA QC, (ii) discuss moving average (MA) optimization methods, (iii) gain insight into practical aspects related to applying MA in daily practice and (iv) describe future prospects to enable more widespread acceptance and application of MA QC. Each of the MA QC optimization methods currently available has their own advantages and disadvantages. Recently developed simulation methods provide realistic error detecting properties for MA QC and are available for laboratories. Operational MA management issues have been identified that allow developers of MA software to upgrade their packages to support optimal MA QC application and guide laboratories on MA management issues, such as MA alarm workup. The new insights into MA QC characteristics and operational issues, together with supporting online tools, may promote more widespread acceptance and application of MA QC.

Keywords: analytical quality control; average of normal; moving average; quality assurance; quality control.

Background

Moving average quality control (MA QC) was first described as "average of normals" by Hoffmann and Waid in 1965 as an analytical quality control (QC) instrument [1]. They proposed a method that averages the results obtained within (more or less) the reference range and plotted these in a control chart. Since then, studies on this average of normals concept have resulted in (i) alternative methods and new algorithms to calculate average values, (ii) deeper understanding of moving average (MA) error detection and its characteristics and (iii) guidance on how to obtain optimal MA settings. Supported by these studies, the average of normals method has evolved into a more general MA approach, which is not necessarily based on "normals" and mean calculations. Together with other QC instruments, e.g. internal (statistical) QC, confirmation and authorization procedures, etc., MA is one of the tools currently available to medical laboratories for QC purposes. The potential of MA QC as a valuable QC instrument has been shown and experts consider MA QC to be a valuable tool to support the analytical quality assurance of at least a selection of the standard biomarkers available in medical laboratories [2–4]. However, despite its potential and efforts, MA QC is still struggling to meet these expectations.

MA suffers from the complexity of obtaining optimal MA settings and the absence of gaining objective insight into its error detection properties. MA optimization methods are generally complex and based on advanced statistics that are not always easily understandable. Also, evidence-based guidance on how and when to use MA QC is often lacking or is not bundled together. To my knowledge, no comprehensive review is available that also provides a critical evaluation of the fundamental concepts and practical aspects of MA QC.

Therefore, the aim of this review is to (i) explain the basic concepts of MA QC, (ii) discuss MA optimization methods, (iii) gain insight into practical aspects related to applying MA in daily practice and (iv) describe future prospects that may promote more widespread acceptance and application of MA QC.

MA QC procedures

MA QC is a mathematical procedure that averages obtained assay results and uses the obtained average values for QC

***Corresponding author: Huub H. van Rossum,** Department of Laboratory Medicine, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066CX, The Netherlands; and Huvaros, The Netherlands, Phone: +31-20-5122756, Fax: +31-20-5122799, E-mail: h.v.rossum@nki.nl

purposes. An MA procedure generally consists of three features: (1) criteria for including assay results in the calculation procedure, (2) a calculation algorithm used to calculate average values and (3) control limits that determine out-of-control alarming.

Inclusion criteria and truncation limits

The inclusion criteria, or perhaps more correctly the exclusion criteria, aim to remove outliers and extremes from the calculation algorithm to reduce variation in the included assay results. Reducing the variation in assay results reduces variation in MA values and application of more stringent control limits, thereby allowing the detection of smaller systematic errors. Theoretically, the inclusion criteria can be based on any variable considered relevant, e.g. the hospital department [5], in/outpatient populations [2], week/weekend day [3], etc. However, most often and generally supported by MA management software packages, these criteria are based on the exclusion of extreme or not-normal assay results. The limits used for this purpose are also referred to as truncation limits [2, 6]; however, this approach has some limitations. For example, the exclusion of too many results from the MA calculations implies that an MA value is less frequently calculated and that MA alarming might be delayed. Also, when truncation limits are used, extreme systematic errors might become undetectable due to the exclusion of all generated systematic errors containing assay results [6]; this is illustrated in the bias detection curves presented in Figure 1B, C and D. To select optimal inclusion criteria, these effects need to be taken into consideration.

Calculation algorithm

The next step is to calculate average values. For this, several algorithms can be used, including the mean [2–4], the median [4], the exponentially weighted moving average (EWMA) and the XbarB [8]. Each calculation algorithm has its own specific features and the use of one or the other can be based on these features. Nowadays, the availability or absence of these algorithms in MA management software packages for laboratories can be another critical factor in selecting a calculation algorithm [3, 9]. From a statistical point of view, the mean calculation is considered suitable for MA with assay results that are

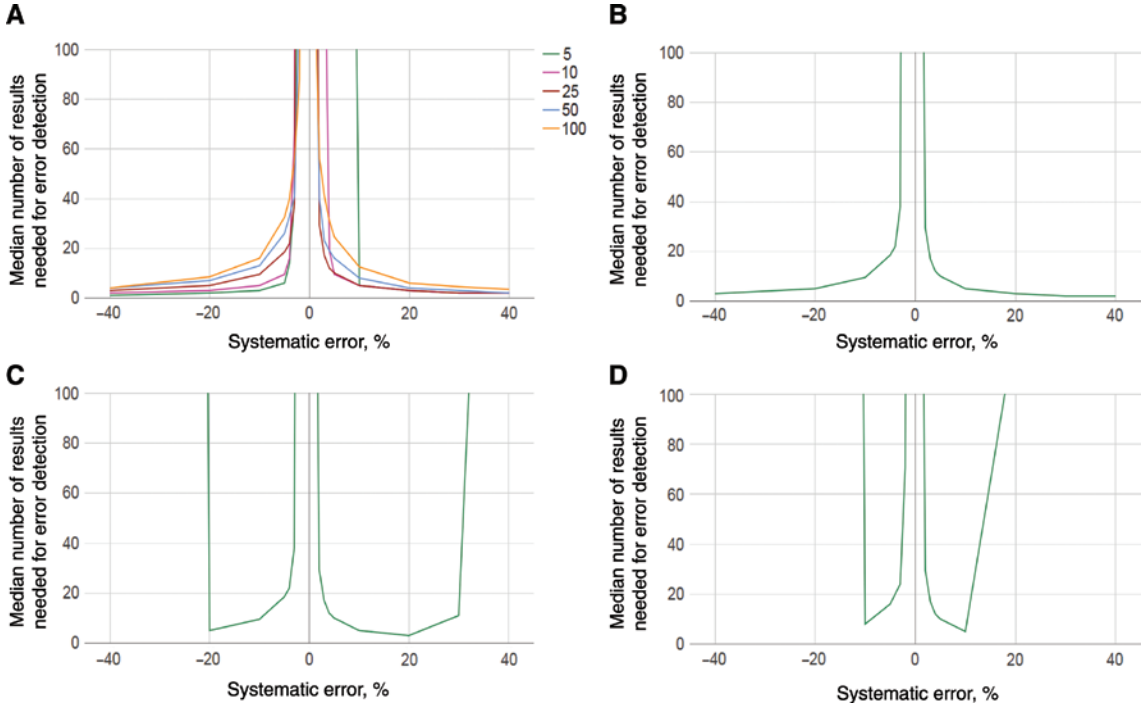

Figure 1: Bias detection curves with and without truncation limits.
Bias detection curves are presented for sodium moving average (MA) procedures. (A) Bias detection performance of 5 MA procedures using mean calculation algorithm with different batch sizes. Sodium MA using mean calculation of the last 25 results is presented without truncation limits, (B) using truncation that only includes results between 100 and 180 mmol/L (C) and only includes sodium results between 120 and 160 mmol/L. (D) Graphs were obtained using the MA Generator application [7].

normally distributed and have a low variation in results, whereas median calculations are generally more powerful for assays with non-normally distributed assay results and those with more extreme results or outliers [10]. The XbarB algorithm described by Bull et al. [8], and also referred to as Bulls algorithm, was designed to smooth the moving average deviation in order to minimize the effect of outliers and allow application for non-Gaussian populations. By design, the XbarB works with batches, and new MA values are generated only after filling up a batch, with a potentially unnecessary delay in bias detection [8, 9]. Alternatively, by design, the EWMA calculates a new MA value for each newly available assay result and, therefore, represents a more true and continuous "moving" average. For median and mean calculations, in terms of filling a batch or "walking" batches, both designs are available. Alternative measures or algorithms include exponentially adjusted moving mean [11], Bhattacharya calculation [12], moving standard deviation and moving sum of outliers [13], number of positive patient results [14] and the average of delta [15]. However, no clear performance-based guidance is available concerning which specific MA algorithm should best be used.

The second part of the calculation algorithm is the selection of variables used in the calculation algorithm. For the mean, median and XbarB, a batch size needs to be selected, whereas for the EWMA, a weighting factor between 0 and 1 is required. Based on the results of Bull et al. [8], an XbarB with a batch size of 20 results is generally used as a standard for the MA procedure for erythrocyte indices and other hemocytometer analytes. In general, with the use of large batch sizes (or a low weighing factor), less variation in MA values is observed and smaller systematic errors can be detected. The trade-off is that the detection of larger systematic errors requires more assay results and can be delayed [5, 6]. A calculation feature that is considered necessary to meaningfully interpret the obtained average values in an MA control chart is that all MA values are computed using the same number of included results and are not, for example, based on a daily average; this avoids MA variation related to the use of different calculation procedures [16].

Control limits

The control limits are assigned MA values that, when exceeded, indicate an out-of-control status of the assay. These limits are generally based on population-based SD rules [4, 17], the reference change value [2] or other related statistical approaches [18]. When using these approaches, the upper and lower control limits are related to each other and have the same distance from the MA mean. More recently, control limits were set by the primary requirement of a manageable or acceptable number of false MA alarms [4, 6]. For this purpose, after running the MA procedure on a training set obtained from the laboratory, the observed minimum and maximum MA values were used as control limits [6]. Using this approach, non-symmetrical and non-SD-based control limits are obtained to meet the requirement of no false MA alarms while using the most stringent control limits [6].

MA optimization methods

For MA application, the first step is to find the optimal MA settings. Probably the most commonly used method to obtain "optimal" MA settings is the trial-and-error approach; this does not require any supportive software, statistics or extensive data analysis. This approach is generally based on the requirement of not accepting too many MA alarms and, based on this requirement, the inclusion criteria, the calculation algorithm and the control limits are selected. However, the use of this method means that no insight is acquired into the bias detection properties of the MA and whether (or not) a truly optimal MA procedure has been applied. Several other approaches are discussed below and are summarized in Table 1.

Initially, the more advanced methods used to generate optimal MA settings were mainly based on graphical interpretation of the MA control graph [8]. This approach was used by Bull et al. [8] to examine various MA algorithms to investigate optimal MA algorithms for controlling the red cell indices. The authors focused on detecting a 3% systematic error and graphically interpreted the MA pattern to select optimal MA algorithms. They concluded that the currently widely applied XbarB was the most optimal algorithm; however, the authors emphasized that their results were not based on an analytic criterion. More recently, a software-supported comparable approach was described that supported the automated MA selection based on the detection of the total error allowable (TEa) and false rejection performance criteria [4].

Power function analysis

Others have used power function analysis to study MA characteristics and to obtain optimal MA settings for routine practice [17, 19, 21]. Power function analysis was

Table 1: Moving average quality control (MA QC) optimization methods.

MA optimization method	Method description	Advantages	Limitations	Ref. no.
Trial and error	MA settings are tried out and based on MA alarming/experience and MA graph, adjustments are made to MA settings	– Simple – Generally available – Based on true patient data	– No insights into systematic error detection properties – No insights into whether optimal MA is obtained	
Graphical	Simulating MA bias detection by plotting error containing MA values in MA control chart	– Direct insights into MA pattern – Based on true patient data	– No analytical performance criteria used – No insights into systematic error detection properties – Based on simulation of 1 systematic error	[8]
Power function analysis	The MA bias detection probability is determined by estimating the % of MA values outside the MA control levels after introduction of systematic error	– Multiple systematic errors are studied – Published nomograms available	– Based on computer-generated laboratory data – Does not properly reflect the effect of truncation limits – Does not take the onset of systematic error in MA values into account	[17, 19]
TEa detection probability	TEa error detection was simulated by adding TEa to obtained patient results and MA procedures with a >90% probability of error detection were selected	– Based on true patient data – Insights into TEa error detection – Included false rejection requirement	– Based on simulation of 1 systematic error – Simulation software requirement	[4]
ANPed	MA error detection is simulated using the number of assay results needed for systematic error detection as readout. Average values of obtained results are used for optimization and validation	– Reflects truncation effects on MA performance – Incorporates the onset of systematic error	– Based on randomized laboratory data – Optimization solely based on TEa – Simulation software requirement	[2, 20]
Bias detection curves and MA validation charts	MA error detection is simulated using the number of assay results needed for systematic error detection as readout. MA performance is presented as median number of assay results (bias detection curves) or number of assay results required to detect error with a 90%, 95% or 100% probability (MA validation chart)	– Based on true patient data – Reflects truncation effects on MA performance – Incorporates the onset of systematic error – Results are risk-based (probability of error detection)	– No clear optimization criteria – Simulation software requirement	[5, 6]

first described and used to thoroughly investigate the performance of statistical QC rules for internal QC results [22, 23]. In line with this application, power function analysis was then used to assess the performance of MA QC procedures [19]. Although highly successful for the analysis of statistical QC performance, this approach has some limitations when used to investigate MA QC. The first limitation is that the analysis is based on computer-generated data with a Gaussian distribution [19]. Although this seems to work when describing internal QC results, this does not apply for most measurands in medical laboratories that often show a severely skewed distribution [2, 6]. Also, the samples and, thereby, the distribution of assay results are generally not randomly presented to the laboratory. They are usually determined by logistics and patient care settings, such as clinical rounds (intensive care, nephrology, pediatrics, etc.), week and weekend days, in- and outpatient programs, etc.; however, all of these have a significant effect on the normal MA pattern or "MA fingerprint" of a laboratory [2, 6]. Another limitation of power function analysis is that the true effect of truncation limits on potentially delaying error detection is not reflected [5]. When truncation limits are applied, large systematic errors can result in the exclusion of many results from the calculation algorithm, thereby significantly delaying error detection [5, 6]. Last but not least, the onset of error in the MA values is not reflected in power function analysis. A larger batch size always results in better error detection [5, 19]; although this is basically true, at the same time a larger batch size always delays the detection of (especially) larger errors [5, 6]. The latter becomes more relevant when MA QC is used for continuous and real-time

QC purposes. An advantage of this approach is that nomograms are available for implementation and represent a more universal and less laboratory-specific optimization method [19].

Simulation methods using results needed for systematic error detection

More recent MA optimization approaches use advanced simulation strategies in which systematic error is introduced, and the number of assay results needed for systematic error detection is used as read-out or output [2, 6, 20]. These approaches aim to simulate the true MA performance, as observed in daily practice. These simulation results can be presented in either the mean number of results needed for systematic error detection as is reflected in the ANPed parameter [2, 20], or the median number of results needed to detect a systematic error [6]. When the simulations are performed for various systematic errors, the results can be plotted in specifically designed bias detection curve graphs in which, for multiple systematic error, the median number of assay results needed for error detection is presented. When multiple MA procedures are plotted, their bias detection performance can be compared (Figure 1A).

Additional information is needed to be able to properly interpret studies using these approaches. Because the results are based on a reference data set, this set has to be representative for a true, in-control, run. Therefore, the method of interest should have been in-control during the reflected data-sampling period. Second, the order of the obtained results should be maintained during the simulations; this is necessary to be representative of the MA fingerprint as would be observed reflecting, for example, week and weekend profiles, in- and outpatients logistics, intensive care, pediatric blood sampling rounds, etc. [2, 6]. Alternatives that apply random sampling from the data set or computer-generated assay results can only estimate a "normal" MA pattern. They do not take into account the true logistic variables that determine the true MA fingerprint [20]. Furthermore, the use of either a mean or median value to interpret the simulation results has different meanings. The median has statistical power in the sense that in 50% of the simulations the systematic error is detected in less than the represented number, and in 50% of the simulations more results were required. For this reason, the median value has an understandable meaning and can be used as a risk parameter [6], whereas the mean value does not provide this kind of information because most of the obtained simulation results are not normally distributed (Figure 2). Alternatively, also the 90%, 95% or the minimum/maximum number of results needed for bias detection can be obtained that allow a more risk-based estimate of the MA QC performance. Therefore, this information (as presented in MA validation charts) can be used to design more risk-based MA QC procedures (Figure 2) [6].

A drawback of these advanced simulation strategies is that the outcomes are probably rather laboratory specific. Furthermore, advanced statistical modeling and/or software is required that is generally not available in medical laboratories, although the web-based application MA Generator has become available online [7]. This application allows laboratories to use the MA optimization approach described by van Rossum et al. [6], using laboratory-specific data sets.

Optimal MA requirements

The last point of discussion concerning MA optimization methods is what parameter to use to guide the MA optimization process. Generally, MA optimization focuses on the detection of a specified systematic error, often the TEa [2, 4, 18]. The risk of using this approach is that by focusing

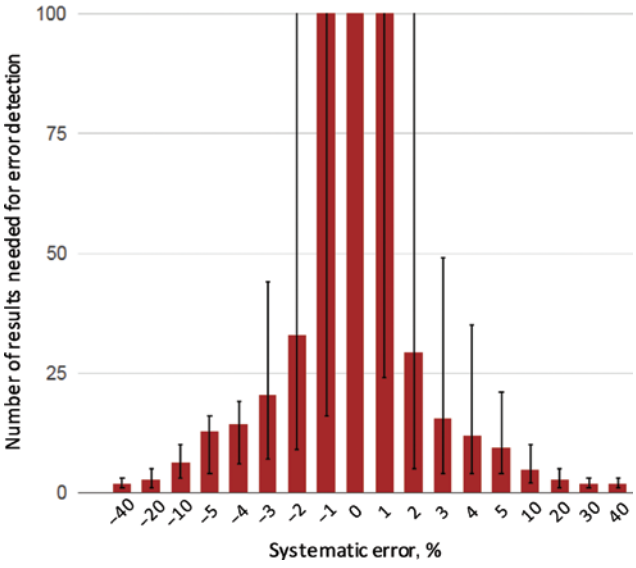

Figure 2: Moving average (MA) validation chart.
A MA validation chart is presented for a sodium MA procedure without truncation limits, calculating the mean of the last 25 results and using 137.5 and 143.6 mmol/L as lower and upper control limit, respectively. Bars represent median number of results, and error bars represent minimum and maximum number of results needed for error detection as observed in the simulation analysis. The graph was obtained using the MA Generator application [7].

on this specific error, the detection of systematic errors larger than the TEa is potentially compromised, e.g. by the use of truncation limits (as shown in Figure 1). Conceptually, TEa or related performance criteria reflect a threshold that determines a clinically relevant error, and therefore, all larger errors are of at least equal importance [24, 25]. In this respect, MA QC differs from statistical QC in the sense that larger errors will always be detected with a larger probability; for MA QC, this depends on the use of truncation limits in an MA procedure and cannot be assumed. This is illustrated in Figure 1. Here the use of truncation limits in Figure 1C and D compromises the detection of larger systematic errors. The MA presented in Figure 1D is unable to detect a 20% systematic error, whereas a systematic error of 10% is rapidly detected. An alternative approach is to use the bias detection curve itself as an optimization parameter. This optimization approach is less clearly guided but allows to take into account the overall systematic error detection performance of MA procedures [6].

Application of MA QC

When considering MA QC as a QC instrument, a first consideration is the purpose of the MA QC. The two main reasons to use MA QC are (i) for assurance of long-term method stability [26–28] and (ii) for continuous/real-time QC [2, 3, 29]. Another MA application is assuring preanalytical stability; examples of the detection of preanalytical errors by MA QC have been described [3]. Also, when no or only QC materials with a limited stability are available, MA QC could be of use.

Long-term assay stability

One application of MA QC is the assurance of long-term assay stability. One reason for this application is the issue of non-commutability that can be associated with internal and external QC samples [30–33]. Examples of using MA approaches as a quality assurance tool include the stability of FT4, TSH, calcium and phosphate using a 50th percentile algorithm. Using a similar approach, clinically relevant shifts in sodium were observed that coincided with reagent lot shifts [34]. Others detected a shift in bone-specific alkaline phosphatase using an MA, which was not visible using external and internal QC [35].

More recently, a more or less pragmatic MA approach was introduced. It comprises two web-based and freely available applications called the Percentiler and Flagger [36, 37]. They are part of a larger empowerment project developed to enhance the communication on analytical performance of commercial test systems between laboratories and in vitro diagnostic manufacturers [26, 27, 31]. The Percentiler database comprises instrument-specific daily medians calculated from outpatient results and sent by e-mail from all over the world. Data are peer grouped by platform/assay. Via a user interface, participants have access to the graphical presentation of the course of their moving medians in comparison with the peer group one. This allows inference of the mid- to long-term stability of analytical performance at the individual and peer group level, and/or occurrence of shifts/drifts [26, 27]. The Flagger application is designed to monitor the stability of the percentage of results flagged when they exceed the cutoff points used by individual laboratories. From combining the Percentiler and Flagger observations, users can infer how the change in flagging rate in their laboratory is related to the analytical variation [38].

Real-time/continuous QC

The second major application of MA QC is continuous (preferably real-time) QC. For the long-term assay stability application of MA QC, generally one MA QC is generated on the longer term, such as on a daily [27], weekly [34] or monthly basis [32]. However, MA QC values can be calculated for every new, or a couple of newly generated assay results, resulting in a more continuous QC process. When the MA QC can detect errors within one or a couple of assay results, e.g. as shown for large systematic errors for sodium MA (Figure 2), it could be claimed that, for these errors, MA QC supports real-time QC. However, by design, erroneous samples have to be generated before MA can detect such an error.

MA QC to support statistical QC

When considering continuous laboratory production processes, bracketed internal QC (iQC) is applied for analytical quality assurance [39]. In current practice with short turn-around times, continuous analysis and the release of multiple diagnostic test results, many laboratory results will be released before a confirmatory QC is performed. When a continuous production process is analytically controlled using scheduled iQC, there is a risk of reporting erroneous results by either rapid onset of (large) error in between the scheduled QC [3] or temporary assay failure in between the scheduled iQC, as demonstrated for a sodium MA alarm case [9].

Furthermore, by design, statistical QC is limited in its ability to detect clinically relevant errors for low sigma processes [39, 40]. Interestingly, these low sigma processes (characterized by a low biological variation/analytical variation ratio) generally have the characteristics for optimal MA performance [19]. Furthermore, when risk-based statistical QC schedules are used for high-volume low sigma (<3) assays, statistical QC becomes almost unmanageable [39]. For these reasons, MA QC can further support and extend the analytical quality assurance of statistical QC [39]. However, to my knowledge, as no study has designed and validated integrated statistical QC and MA QC strategies, there is no clear guidance regarding how to design risk-based QC plans incorporating MA QC. Furthermore, it should be noted that, for several reasons, it is unlikely that MA can replace the measurement of QC materials (internal and/or external) [33]. For example, analytical failure detected by MA should be confirmed, e.g. by QC samples, and MA error detection performance for many measurands is limited. Also, after performing potential error-introducing procedures (such as large maintenance, new reagent lots, calibrations, etc.), QC measurement is required before analyzing samples in order to confirm proper assay performance, which cannot be done by MA QC [33].

Implementation of MA QC

Implementation of MA QC can be divided into the following stages: (i) determination of the MA procedure settings (truncation limits, calculation algorithm, control limits); (ii) configuration of MA settings and MA management on the analyzer, middleware or other MA supporting software packages; and (iii) the design and implementation of laboratory protocols for MA alarms [9]. Methods to optimize and validate MA procedures are described above.

After obtaining optimal MA settings, they are implemented on software packages available in laboratories that support the MA application. In most (if not all) hemocytometer analyzers, some type of MA package is available that generally supports at least the XbarB calculation algorithm [8]. Also, several middleware systems, including Remisol Advance (Beckman Coulter), Instrument manager (Data Innovations), Roche Middleware Solutions (Roche Diagnostics), Centralink Data Management System (Siemens Healthcare Diagnostics) [41], AlinIQ [42], as well as laboratory information systems such as GLIMS [3], support the MA application. In practice, the software packages available in laboratories determine which calculation algorithm is available, as some are restricted to only the mean or XbarB calculations [3, 9], whereas others support a larger range of MA algorithms [43]. Furthermore, these software packages are used for MA management, including real-time MA calculations, graphical presentation of MA, alarming, etc. Until now, few studies have focused on issues related to practical MA application. One prospective trial applied MA for continuous QC in routine practice for 24 chemistry assays run on two analyzers [3]. The study identified several MA software features that were considered necessary or helpful to support the MA application. Furthermore, several issues were identified that complicated MA management in daily practice. Based on these observations, several software features required to optimally support MA management in daily practice are listed in Table 2. Others used software for MA management that was capable of a "release for the back" approach, in which assay results are only released after passing of MA QC [4]. Having software that supports MA management in a proper way is an important aspect when considering MA QC for continuous analytical QC.

Finally, when MA is setup in the MA management software, laboratory protocols for handling MA QC alarms should be designed and implemented. The design of these protocols also determines which errors detected by MA alarms will be acknowledged. For example, a case was reported in which sodium assay failure was detected by MA [9]. However, at the time of QC measurement as part of the MA alarm workup, there was no sign of assay failure. Only reanalysis of the patient samples as part of the MA alarm workup was able to elucidate that the MA alarm detected temporary, clinically significant, assay failure [9]. Because MA alarms can be due to causes other than analytical assay failure (e.g. preanalytical issues, a single patient with extreme results, or false MA alarms), follow-up of an MA alarm should (when possible) at least include internal QC analysis in order to confirm analytical error [3, 9].

Outlook

MA QC has been around for decades and remains a promising QC instrument. Nevertheless, more widespread application of MA is hindered by (i) the complexity of establishing optimal MA settings, (ii) a lack of evidence-based guidance on how to use MA and (iii) clear evidence that supports its added value. Although many believe in the value of MA application, they lack the resources and/or the time and/or expertise to implement it. Others

Table 2: Desirable moving average (MA) management software features [3, 9].

Subject	Feature	Comment
MA inclusion and exclusion criteria	– Truncation limit application	For many MA procedures truncation limits are required
	– Support exclusion of non-patient samples In order to avoid false MA caused by; iQC, eQC, research specimens, dialysis fluids, etc	
	– Exclusion of selected patients	Supports MA management; avoids future false MA-alarms if MA-alarm is triggered by a single patient with extreme results
Calculation algorithms	– Supports the mean, median, EWMA and XbarB algorithms	These are the most commonly used algorithms that all have rather specific indications when to be used
MA calculation frequency	– Supports MA calculation for every new assay results	This continuous calculation supports real-time MA QC, without unnecessary delay of systematic error detection
Control limits	– Support of SD-based as well as accuracy or MA reference range, based control levels	Depends on the MA optimization method used
MA Graphical presentation	Support MA presentation in Levey-Jennings or Accuracy plot (Figure 3)	Historically MA have been plotted in Levey-Jennings plots. New approaches use MA reference range which is not based on SD values
	Be able to exclude MA values from iQC Levey-Jennings plot	Because many MA results can be obtained, plotting these in the same graph as the iQC, iQC results plotted in the same graph can be masked
MA reset	Be able to reset an MA after MA-alarm workup	Especially for continuous MA, it can take multiple assay results before an MA is within its control limits after an MA alarm. A forced reset after workup of an MA-alarm avoids unnecessary disturbing MA alarms
MA alarming	Real-time notification of MA-alarm (push set-up)	When MA is used for continuous QC, alarming should be continuous and technicians should be actively notified

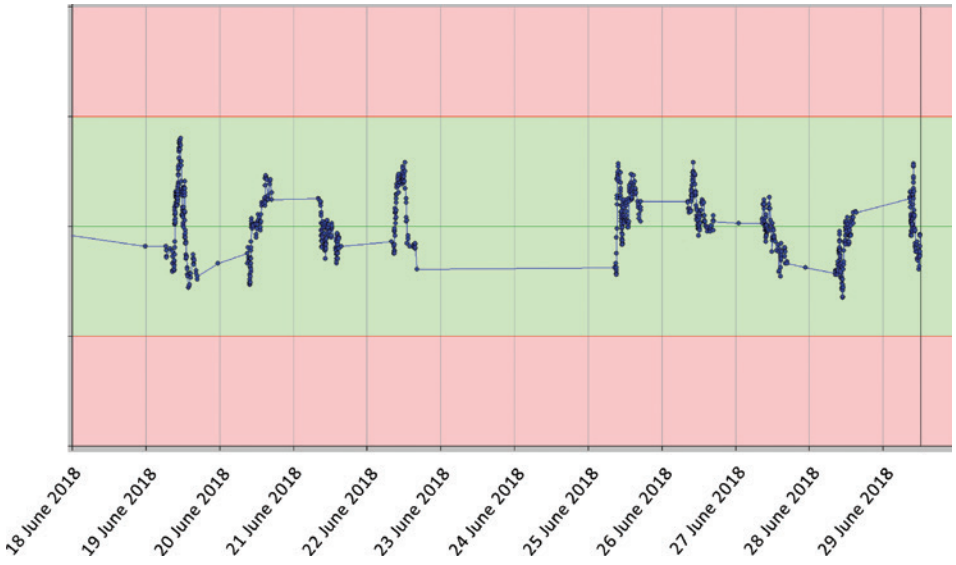

Figure 3: Moving average (MA) graph presented in accuracy plot.
The MA graph presents a sodium MA procedure. The sodium MA procedure is the same as that used for the MA validation chart in Figure 2. The green area shows the MA values between the control limits determined according to van Rossum et al. [5, 6].

have used MA but were initially disappointed or discouraged, perhaps due to improper use and/or too high expectations.

It appears that several steps are required to support more widespread and proper application of MA and allow laboratories to optimally benefit from MA QC. First, it should be acknowledged that MA QC is significantly different from statistical internal QC, i.e. MA QC has different properties (Table 3) and statistical QC validation concepts do not necessarily apply to MA QC. For example, an important difference is reflected in the effect of using truncation limits on compromising the detection of larger bias [5, 6].

Table 3: Characteristics of moving average (MA) and statistical quality control (QC).

Characteristic	Moving average QC	Statistical internal QC
QC frequency	Continuous	Scheduled
Commutability	Commutable	Risk of non-commutability
Controllable diagnostic phases	Preanalytical phase and analytical phase	Analytical phase
QC level	One level dependent on patient population and MA settings	Multiple adjustable levels
Error detection	Bias (imprecision unknown)	Bias and imprecision
Optimization and validation	See Table 1	Statistics (SD)/sigma metrics/risk-based
Graphical presentation of results	Accuracy plot/Levey-Jennings plot	Levey-Jennings plot
Operational costs	MA-QC alarm workup	QC materials, QC analysis and QC alarm work-up

Nevertheless, these different characteristics can be seen as an opportunity to strengthen QC plans by using MA QC and making use of the best quality assurance characteristics of both statistical QC and MA QC. As mentioned, MA can potentially support the low sigma analytical processes that are often difficult to control with statistical QC, support real-time QC and alert for preanalytical instabilities. However, to achieve all this, there is a need for (new) ways to design and validate QC plans based on integrated statistical QC and MA QC.

Another issue is the availability of MA optimization methods and proper MA management software for laboratories. This first essential requirement has at least partly been resolved by the online MA generator [7]. The latter still requires attention; theoretically, MA is supported by many software packages commercially available, and occasionally, limited features are available to support optimal management of real-time MA QC (Table 2). Moreover, a limited selection of calculation algorithms is often available, e.g. MA in hemocytometry is focused on the XbarB, whereas this selection was not based on objectified MA performance criteria. In this respect, future research should also examine the performance of different calculation algorithms using the newly described optimization and validation methods to allow evidence-based use of the most powerful MA algorithms [2, 6, 20].

Summary

Important steps have been made in elucidating the characteristics, potential applications and limitations of MA QC. Representative and realistic MA optimization and validation methods have been developed, and MA optimization methods and MA application for external quality assessment are now available for laboratories. Tools are now available to take MA to the next level and resolve some outstanding issues, such as the integration of MA QC and statistical QC, and appropriate guidance on selecting the most powerful MA algorithms. Practical MA management issues have been identified that allow developers of MA management software to upgrade their packages to optimally support the MA QC application. Thus, although MA has been around for decades, MA QC has reached a tipping point and seems poised for more widespread acceptance and application.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: 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.

References

1. Hoffmann RG, Waid ME. The “average of normals” method of quality control. Am J Clin Pathol 1965;43:134–41.
2. 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.
3. 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;55:1142–51.
4. Fleming JK, Katayev A. Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: for patients by patients. Clin Biochem 2015;48:508–13.
5. van Rossum HH, Kemperman H. A method for optimization and validation of moving average as continuous analytical quality control instrument demonstrated for creatinine. Clin Chim Acta 2016;457:1–7.
6. van Rossum HH, Kemperman H. Optimization and validation of moving average quality control procedures using bias detection

curves and moving average validation charts. Clin Chem Lab Med 2017;55:218–24.
7. MA Generator. Available at: www.huvaros.com. Accessed: June 2018.
8. 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.
9. van Rossum HH, Kemperman H. Moving average for continuous quality control: time to move to implementation in daily practice? Clin Chem 2017;63:1041–3.
10. Wilson A, Roberts WL, Pavlov I, Fontenot J, Jackson B. Patient result median monitoring for clinical laboratory quality control. Clin Chim Acta 2011;412:1441–6.
11. Smith FA, Kroft SH. Exponentially adjusted moving mean procedure for quality control. An optimized patient sample control procedure. Am J Clin Pathol 1996;105:44–51.
12. Oosterhuis WP, Modderman TA, Dinkelaar RB, Zwinderman AH, van der Helm HJ. Bhattacharya: a new application for quality control. Ann Clin Biochem 1991;28:386–92.
13. 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;52:112–6.
14. 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;55:1709–14.
15. Jones GR. Average of delta: a new quality control tool for clinical laboratories. Ann Clin Biochem 2016;53:133–40.
16. Reed AH. Use of patient data for quality control of clinical laboratory tests. Clin Chem 1970;16:129–34.
17. Carey RN. Patient population controls. Clin Lab Med 2013;33:139–46.
18. Liu J, Tan CH, Loh TP, Badrick T. Verification of out-of-control situations detected by "average of normal" approach. Clin Biochem 2016;49:1248–53.
19. 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.
20. Ye JJ, Ingels SC, Parvin CA. Performance evaluation and planning for patient-based quality control procedures. Am J Clin Pathol 2000;113:240–8.
21. 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;42:1683–8.
22. Westgard JO, Groth T. Power functions for statistical control rules. Clin Chem 1979;25:863–9.
23. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule shewhart chart for quality control in clinical chemistry. Clin Chem 1981;27:493–501.
24. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974;20:825–33.
25. Stockl D, Baadenhuijsen H, Fraser CG, Libeer JC, Petersen PH, Ricos C. Desirable routine analytical goals for quantities assayed in serum. Discussion paper from the members of the external quality assessment (EQA) working group a on analytical goals in laboratory medicine. Eur J Clin Chem Clin Biochem 1995;33:157–69.
26. De Grande LA, Goossens K, Van Uytfanghe K, Stockl D, Thienpont LM. The empower project – a new way of assessing and monitoring test comparability and stability. Clin Chem Lab Med 2015;53:1197–204.
27. Goossens K, Van Uytfanghe K, Twomey PJ, Thienpont LM. Monitoring laboratory data across manufacturers and laboratories – a prerequisite to make "big data" work. Clin Chim Acta 2015;445:12–8.
28. Lott JA, Smith DA, Mitchell LC, Moeschberger ML. Use of medians and "average of normals" of patients' data for assessment of long-term analytical stability. Clin Chem 1996;42:888–92.
29. Bennett ST. Continuous improvement in continuous quality control. Clin Chem 2016;62:1299–301.
30. 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;57:76–83.
31. Van Houcke SK, Stepman HC, Thienpont LM, Fiers T, Stove V, Couck P, et al. Long-term stability of laboratory tests and practical implications for quality management. Clin Chem Lab Med 2013;51:1227–31.
32. Jorgensen LM, Hansen SI, Petersen PH, Soletormos G. Median of patient results as a tool for assessment of analytical stability. Clin Chim Acta 2015;446:186–91.
33. Badrick T, Graham P. Can a combination of average of normals and "real time" external quality assurance replace internal quality control? Clin Chem Lab Med 2018;56:549–53.
34. Stepman HC, Stockl D, Stove V, Fiers T, Couck P, Gorus F, et al. Long-term stability of clinical laboratory data: sodium as benchmark. Clin Chem 2011;57:1616–7.
35. 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;54:403–6.
36. The Percentiler. Available at: https://www.thepercentiler.be/. Accessed: May 2018.
37. The Flagger. Available at: https://www.theflagger.be/. Accessed: May 2018.
38. Thienpont LM, Stöckl D. Percentiler and flagger – low-cost, on-line monitoring of laboratory and manufacturer data and significant surplus to current external quality assessment. J Lab Med June 14, 2018 as doi.org/10.1515/labmed-2018-0030. [Epub ahead of print].
39. Westgard JO, Bayat H, Westgard SA. Planning risk-based sqc schedules for bracketed operation of continuous production analyzers. Clin Chem 2018;64:289–96.
40. Schoenmakers CH, Naus AJ, Vermeer HJ, van Loon D, Steen G. Practical application of sigma metrics qc procedures in clinical chemistry. Clin Chem Lab Med 2011;49:1837–43.
41. Middleware systems, cap today. Available at: http://www.captodayonline.com/2014/ProductGuides/06-14_CAPTODAY_Middleware.pdf. Accessed: June 2018.
42. Aliniq analyzer management system. Abbott. Available at: https://www.corelaboratory.abbott/int/en/offerings/brands/AlinIQ/aliniq-ams-features. Accessed: May 2018.
43. Instrument manager. Data innovations. Available at: https://datainnovations.com/news/7519. Accessed: May 2018.