



Ashley D. Ellis, Alexander R. Gross, Jeffrey R. Budd and W. Greg Miller\*

# Influence of reagent lots and multiple measuring systems on estimating the coefficient of variation from quality control data; implications for uncertainty estimation and interpretation of QC results

<https://doi.org/10.1515/cclm-2020-0320>

Received March 15, 2020; accepted March 24, 2020

## Abstract

**Background:** Clinical laboratories use internal quality control (QC) data to calculate standard deviation (SD) and coefficient of variation (CV) to estimate uncertainty of results and to interpret QC results. We examined the influence of different instruments, and QC and reagent lots on the CV calculated from QC data.

**Methods:** Results for BioRad Multiquel frozen liquid QC samples over a 2-year interval were partitioned by QC and reagent lots. The mean and CV were calculated for each partition for each of three Abbott Architect c8000 instruments for measuring serum alanine amino transferase (ALT), creatinine (enzymatic), glucose and sodium.

**Results:** CVs differed among partitions and instruments for two QC levels by 5.8- and 3.3-fold for ALT, by 4.7- and 2.1-fold for creatinine, by 2.0- and 2.6-fold for glucose, and by 2.1- and 2.0-fold for sodium. Pooled CVs for two QC levels varied among instruments by 1.78- and 1.11-fold for ALT, by 1.63- and 1.11-fold for creatinine, by 1.08- and 1.06-fold for glucose, and by 1.24- and 1.31-fold for sodium.

**Conclusions:** The CVs from QC data varied substantially among QC and reagent lots and for different identical specification instruments. The CV used to estimate uncertainty for a measurement result or as the basis for interpreting individual QC results must be derived over a sufficient time interval to obtain a pooled CV that represents “typical” performance of a measuring system. An estimate of uncertainty provided to users of laboratory results will itself have uncertainty that can influence medical decisions.

**Keywords:** internal quality control; quality control; uncertainty.

## Introduction

All measurements in laboratory medicine have an associated uncertainty. Uncertainty represents the influence of sources of variability in a measurement result. Uncertainty (symbol  $u$ ) is expressed as standard deviation (SD), or relative SD as coefficient of variation (CV) in percent. The estimate  $u$  is multiplied by a coverage factor of 2 to determine the expanded uncertainty (symbol  $U$ ) whose positive to negative interval statistically represents a 95.45% probability that the true value is within those limits. The expanded uncertainty is informative for medical decisions because the true value of the measurand is within the uncertainty approximately 95% of the time. All sources of variability in the measurement process contribute to the uncertainty. Sources of variability such as measurand stability during collection, transportation and storage prior to the measurement process also contribute to uncertainty of the final value from the perspective of clinical interpretation [1]. However, the pre-measurement process variability is typically not included when estimating the uncertainty of the measurement process because it is difficult to determine and can vary with both clinical situations and pre-examination conditions. Consequently, the pre-measurement process variability can be different for different measurements using the same measuring system.

The International Organization for Standardization (ISO) standard 15189 specifies requirements for quality and competence in medical laboratories [2]. ISO 15189 is widely used as the basis for accrediting medical laboratories and requires that measurement uncertainty be determined and that the uncertainty be made available on request to laboratory users. ISO 15189 does not require that uncertainty be reported along with laboratory results. Note 1 in ISO 15189 clause 5.5.1.4 clarifies that the relevant uncertainty components are those associated with the actual measurement process, commencing with the

\*Corresponding author: W. Greg Miller, Department of Pathology, Virginia Commonwealth University, PO Box 980286, Richmond, VA 23298-0286, USA, Phone: +1.804.828.0375, Fax: +1.804.828.0353, E-mail: greg.miller@vcuhealth.org

Ashley D. Ellis and Alexander R. Gross: Department of Pathology, Virginia Commonwealth University, Richmond, VA, USA

Jeffrey R. Budd: Jeff Budd Consulting, St. Paul, MN, USA

Open Access. ©2020 Ashley D. Ellis et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

presentation of the sample to the measurement procedure and ending with the output of the measured value.

There are two approaches to estimate uncertainty of a laboratory measurement described in the *Guide to Uncertainty of Measurement* [3]. A bottom-up approach individually estimates each source of variability in a measurement process and combines them to estimate the uncertainty. A bottom-up approach is useful when developing a measurement procedure to identify dominant sources of variability that can be reduced to meet the combined uncertainty goal. A bottom-up approach is difficult to apply for a clinical laboratory because information on uncertainty for many of the individual steps in a measuring system is not available. A top-down approach estimates the combined uncertainty from replicate measurement results that reflect all or most sources of variability in a measuring system, for example, from quality control (QC) results [4]. The Clinical and Laboratory Standards Institute document EP29 provides guidance on estimating uncertainty using either the bottom-up or the top-down approach [5]. ISO technical specification, TS 20914, for estimating uncertainty in medical laboratories uses the top-down approach based on QC data [6]. For a clinical laboratory, the top-down approach using data from QC results is practical because the data are readily available. Uncertainty based on QC data can be combined with the uncertainty of values assigned to end-user calibrators to estimate the combined uncertainty of results from a measuring system used in a clinical laboratory.

We examined the variability of CVs estimated from QC data for use to determine a “typical” uncertainty for reported results or for use to interpret individual QC results to determine if a measuring system is operating within its performance specifications. In particular, we examined the influences of QC lot and reagent lot on the estimate of CV over short and long time intervals. We also investigated the situation when a patient’s sample has a random chance to be measured on one of three identical specification measuring systems.

## Materials and methods

Three Abbott Diagnostics c8000 measuring systems for serum alanine amino transferase (ALT), creatinine (enzymatic), glucose and sodium were operated according to the manufacturer’s instructions for use with reagents and calibrators provided by Abbott. The three c8000 instruments were part of an automated track sample delivery system with a random chance that a given patient’s sample would be measured by one of the three c8000 instruments.

Internal QC samples were Multiqual frozen liquid unassayed levels 1 and 3 (referred to as level 2 in this report) from BioRad Life Sciences. Two lots of each QC level were used as well as six to nine

lots of reagents for each analyte during the 2-year interval of this investigation as indicated in Supplementary Table S1. QC results were discarded and not used if those results identified unacceptable performance that required corrective action and repeat measurement of patients’ samples, or were blunders when results for the two levels were mixed-up.

QC results were partitioned by QC lot and by reagent lot within each QC lot as shown in Supplementary Figure 1 and Supplementary Table S1. The mean and CV were calculated for each partition for each of the three measuring systems. A pooled CV was calculated for each QC level for each measuring system using the following equation [5, 6].

$$CV_{\text{pooled QC}} = \sqrt{\frac{(n_1 - 1) \times CV_1^2 + (n_2 - 1) \times CV_2^2 + \dots + (n_i - 1) \times CV_i^2}{n_1 + n_2 + \dots + n_i - i}}$$

where  $n_i$  is the number of observations in a partition and  $CV_i$  is the CV for QC data in a partition. CVs were used in the data presentation because the mean values differed among different QC and reagent lots making SDs more difficult to compare. Confidence intervals (CIs) for CVs pooled across QC and reagent lots for each instrument were estimated using restricted maximum likelihood (REML) variance component estimates (JMP software, SAS Institute). REML is suitable for unbalanced data with different numbers of individual QC results for different reagent lots and instruments. REML gave residual CVs identical to the pooling equation CVs plus CI values. Overall total CVs and CIs across analyzers were calculated using REML which included between- and within-instrument components.

Each c8000 instrument measured, in singlet, each analyte from a pooled patient sample (PS) prepared each week during the 2-year interval. A PS consisted of serum or heparin plasma from three to five patients stored capped at 2–8 °C for 2–3 days before pooling. PSs were prepared and measured on the same day. PSs had different concentrations on different weeks intended to examine consistency of results for a reasonable portion of the measuring interval over time. The mean and CV were calculated for weekly results for each PS from the three c8000 instruments. For sodium, as CV pooling over all weeks gave an estimate close to the median CV over all weeks, the distribution of  $CV^2$  was unskewed and was thus used to estimate CV and its CI. ALT results were transformed as  $\log(X + 0.2)$  and creatinine results as  $\log(X + 0.3)$  to achieve a normal distribution of CV results. For glucose, no transformations were found to improve the distribution of CV over all weeks and glucose data was not transformed. The Anderson-Darling test of normality informed the decisions for ALT, creatinine and glucose. For these three measurands, the CI for the average CV was calculated from the SD for the normal or transformed distributions of weekly CVs. Outliers eliminated from the PS data are explained in the Supplementary data.

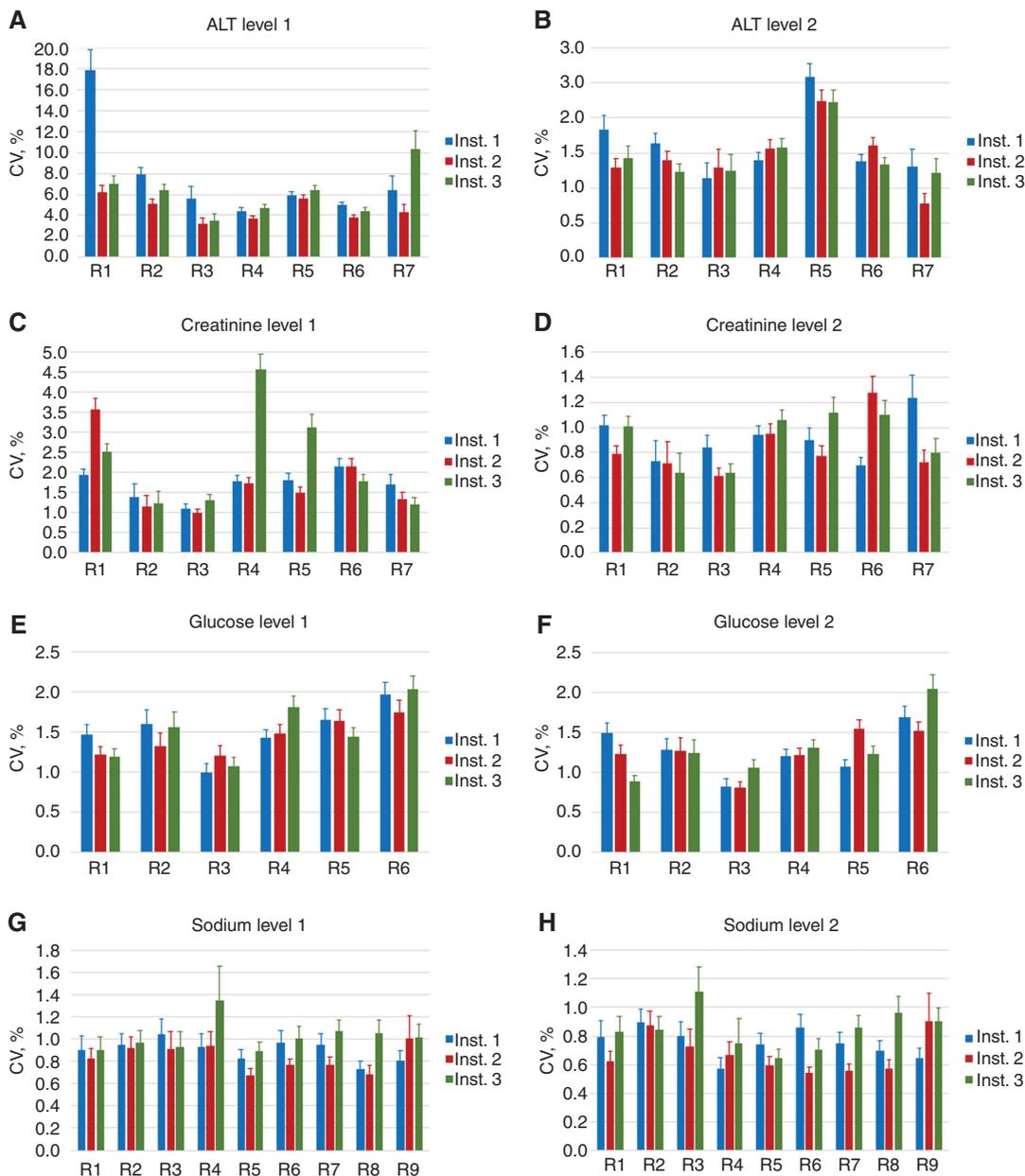
## Results

The statistical summary of QC results for each partition for each analyte are shown in Supplementary Table S1 along with the pooled CVs for QC data and the average CVs for the PS measured weekly by each instrument for each analyte.

ALT

Figure 1A and B show CVs for individual partitions for ALT. For QC level 1, CVs among all reagent lots on an individual instrument varied 2.00–4.09-fold among the three instruments for level 1, and 1.85–2.87-fold for level 2. For the same reagent lot, CVs among the three instruments varied 1.15–2.89-fold for QC level 1 and 1.13–1.67-fold for QC level 2. CVs among all combinations of reagent lots and instruments varied 5.77-fold and 3.32-fold for QC levels

1 and 2, respectively. Examination of Levey-Jennings plots of QC data for level 1 instrument #1 in partition R1 and instrument #3 in partition R7 showed small shifts or drifts in values that remained within the acceptance criteria in use by the laboratory but explain the larger CVs for those partitions (Supplementary Figures S2 and S3). Examination of Levey-Jennings plots of QC data for level 2 showed increased imprecision for reagent lot 5 for all instruments but no outlier values (Supplementary Figures S4–S6).



**Figure 1:** CV for each QC and reagent lot partition for each instrument over a 2-year interval. The error bar shows the upper CI for each CV. The panels show results for two levels of QC samples for ALT (A and B), creatinine (C and D), glucose (E and F) and sodium (G and H). The number, mean and range of QC values is shown in Supplementary Table S1.

Figure 2A shows the CV pooled across all QC and reagent lot partitions for ALT for each QC level and for each instrument. Pooled CVs were 1.78- and 1.11-fold different between the three instruments for QC levels 1 and 2, respectively. Also shown in Figure 2A is the pooled CV for all QC data across all instruments with values of 6.44% for level 1 and 1.69% for level 2. Figure 2A also shows that the average CV from weekly measurements of PS was similar to the pooled CVs from the QC samples for level 2 (approximately 180 U/L) but smaller than for level 1 (approximately 15 and 23 U/L for the two QC lots examined). It is difficult to compare the pooled CVs for PSs and for QC samples because of the large differences in activities between them. Supplementary Figure S7 shows that the weekly PSs had ALT activities between 16 and 1387 U/L with median 79 U/L, 80% >40 U/L and 60% between 40 and 200 U/L supporting that the CV for the PS is expected to be closer to the pooled value for QC level 2.

## Creatinine

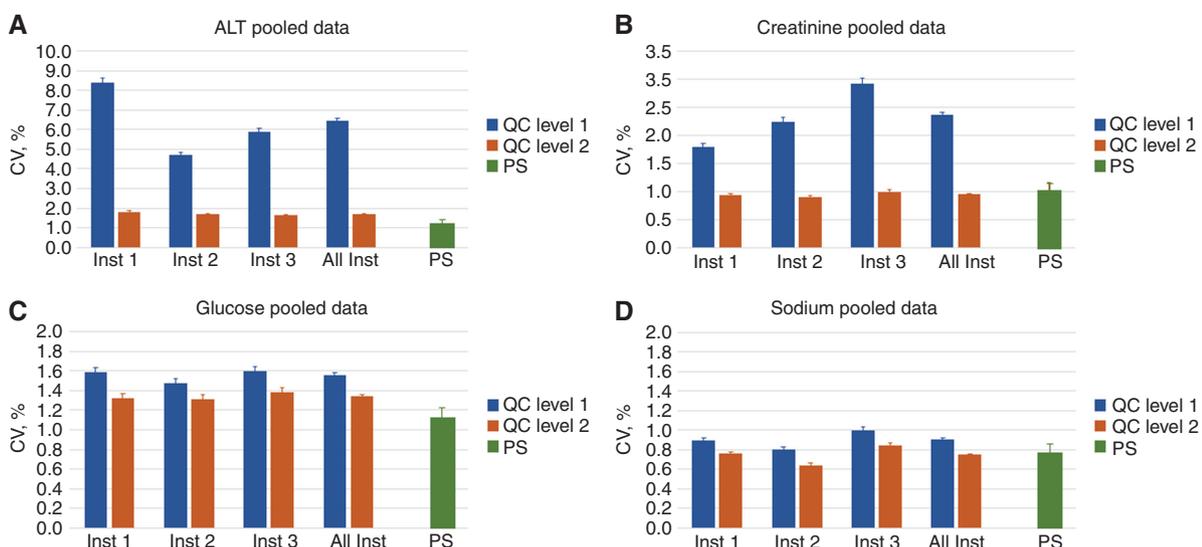
Figure 1C and D show CVs for individual partitions for creatinine. For QC level 1, CVs among all reagent lots on an individual instrument varied 1.96–3.79-fold among the three instruments for level 1, and 1.76–2.09-fold for level 2. For the same reagent lot, CVs among the three instruments varied 1.20–2.65-fold for QC level 1 and 1.13–1.84-fold for QC level 2. CVs among all combinations of reagent lots and

instruments varied 4.70-fold and 2.09-fold for QC levels 1 and 2, respectively. Examination of Levey-Jennings plots of QC data for level 1 instrument #2 in partition R1 and instrument #3 in partitions R4 and R5 showed a small shift or drift in values that remained within the acceptance criteria in use by the laboratory but explains the larger CVs for those partitions (Supplementary Figures S8–S10).

Figure 2B shows the CV pooled across all QC and reagent lot partitions for creatinine for each QC level and for each instrument. Pooled CVs were 1.63- and 1.11-fold different between the three instruments for QC levels 1 and 2, respectively. Also shown in Figure 2B is the pooled CV for all QC data across all instruments with values of 2.36% for level 1 and 0.93% for level 2. Figure 2B also shows that the average CV from weekly measurements of PS was similar to the pooled CVs from the QC samples for level 3 but smaller than for level 1. Supplementary Figure S11 shows that 70% of the PS had creatinine values 106–424  $\mu\text{mol/L}$  (1.2–4.8 mg/dL) supporting that the CV for the PS is expected to be closer to the pooled value for QC level 2.

## Glucose

Figure 1E and F show CVs for individual partitions for glucose. For QC level 1, CVs among all reagent lots on an individual instrument varied 1.45–1.99-fold among the three instruments for level 1, and 1.92–2.33-fold for level 2. For the same reagent lot, CVs among the three instruments



**Figure 2:** CV for pooled QC data from all reagent lots for two levels of QC separately for each instrument, then for all reagent lots and instruments over a 2-year interval.

The green bar is the average CV for one PS measured weekly over the 2-year interval. The error bar shows the upper CI for each CV. The panels show results for ALT (A), creatinine (B), glucose (C) and sodium (D).

varied 1.15–1.27-fold for QC level 1 and 1.02–1.69 -fold for QC level 2. CVs among all combinations of reagent lots and instruments varied 2.05-fold and 2.56-fold for QC levels 1 and 2, respectively.

Figure 2C shows the CV pooled across all QC and reagent lot partitions for glucose for each QC level and for each instrument. Pooled CVs were 1.08- and 1.06-fold different between the three instruments for QC levels 1 and 2, respectively. Also shown in Figure 2C is the pooled CV for all QC data across all instruments with values of 1.56% for level 1 and 1.34% for level 2. Figure 2C also shows that the average CV from weekly measurements of PS was smaller but generally similar to the pooled CVs from the QC samples.

## Sodium

Figure 1G and H show CVs for individual partitions for sodium. For QC level 1, CVs among all electrodes on an individual instrument varied 1.44–1.52-fold for level 1, and 1.56–1.72-fold for level 2. For the same electrode, CVs among the three instruments varied 1.06–1.53-fold for QC level 1 and 1.06–1.69-fold for QC level 2. CVs among all combinations of electrodes and instruments varied 2.06-fold and 2.01-fold for QC levels 1 and 2, respectively.

Figure 2D shows the CV pooled across all QC and reagent lot partitions for sodium for each QC level and for each instrument. Pooled CVs were 1.24- and 1.31-fold different between the three instruments for QC levels 1 and 2, respectively. Also shown in Figure 2D is the pooled CV for all QC data across all instruments with values of 0.90% for level 1 and 0.75% for level 2. Figure 2D also shows that the average CV from weekly measurements of PS was similar to the pooled CVs from the QC samples.

## Discussion

### Estimating variability in a measuring system

Results for the same lot of a QC material are frequently different for different reagent lots measuring the same analyte when results for patient samples are equivalent between those reagent lots [7, 8]. This difference for QC values is caused by different matrix interactions between the non-commutable QC material and the reagent lots that is not present with patient samples. Consequently, QC data must be partitioned by reagent lot to avoid a matrix-related change in values from influencing the mean, SD and CV.

We found, for four representative analytes, that the CV for QC data varied from 1.4- to 4.1-fold between different reagent lots used with a single instrument and 1.1–2.9-fold between three identical specification instruments using the same QC and reagent lots (Figure 1). The CV among all combinations of reagent lots and instruments varied from 2.0- to 5.8-fold. Pooling the CVs from different QC and reagent lot partitions over a 2-year interval gave more consistent estimates of variability of the measurements with CVs that varied from 1.1 to 1.8 fold (Figure 2).

We found the pooled CVs for QC data were generally similar for three identical specification instruments and were also approximately the same as average CV data for PS measured on each instrument for glucose and sodium (Figure 2). For ALT and creatinine, the average CVs for PS were similar to the pooled CVs for QC data with values similar to those of the PS.

In addition to reagent lots, other influences on variability in QC results include calibration and changes in calibrator lots, deterioration of instrument parts and maintenance. Because calibrator lots have the same influence on results for QC and PS, assuming the reagent lot does not change, the influence of calibration and calibrator lot changes is reflected in the QC data [9, 10]. A reasonable assumption is that similar operational characteristics and maintenance cycles occurred in each QC/reagent lot partition, supporting that partitioning QC data by reagent lot and pooling the CVs from partitions gives a representative estimate of the overall CV [9, 10].

We conclude that estimating the variability of a measuring system requires QC data collected over large time intervals to ensure a representative value for SD or CV. These observations have important implications for estimating the uncertainty of an individual measured result and for establishing parameters used to evaluate QC results in daily practice.

### Estimating uncertainty

Laboratories estimate the uncertainty of results to inform users of the confidence they can have interpreting a result relative to decision values or reference intervals. In addition, the uncertainty is useful when interpreting changes in results as reflective of changes in the physiological condition of a patient. The preferred approach for a medical laboratory to estimate uncertainty is as the SD or CV for QC data that represents the combined influence of all sources of variability in the measurement procedure with exception of the uncertainty of the values assigned to calibrators.

Our results show that an estimate of the CV from QC data varied substantially among different reagent lots and among different instruments. Consequently, an estimate of uncertainty made from QC data over a short time interval, for example, a few weeks or months, could be a substantial over- or under-estimate and is an inappropriate approach.

When calculating uncertainty using QC data over a long time interval, the influence of different reagent lots on the QC values must be considered and the data partitioned by reagent lot to calculate mean, SD and CV. Those partitioned values are then pooled into a single estimate of SD or CV that represents the “typical” uncertainty in a result from an individual instrument measuring system. The time interval (number of partitions) for the QC data to give a good estimate of pooled CV depends on the magnitude of the influence quantities that can only be determined by examining a substantial time interval with several changes in reagent lots. When a patient sample might be measured using one of several instruments in a health care system, the added variability among different instruments will increase the uncertainty of an individual result. Consequently, the estimate of pooled SD or CV must include the variability associated with different instruments.

Our results suggest that estimating uncertainty for a test result is difficult and the estimate itself has uncertainty. The pooled CV represented a best available estimate but was still larger or smaller than the actual CV for a particular combination of reagent lot and instrument at a point in time. This observation means that the uncertainty reported by a laboratory to guide medical decisions is a reasonable estimate but not an absolute value, and that medical judgement considering other clinical and physiological information is also important when interpreting laboratory test results. In particular, caution is needed when using a published uncertainty value when making conclusions near decision values or based on reference change values because of the uncertainty in the published uncertainty value [11].

Although not addressed in our data, when there is a substantial difference in uncertainty for a test result, as is frequently observed for point of care devices vs. main laboratory measuring systems, a laboratory should provide separate estimates of uncertainty for the different measurement technologies.

## Establishing parameters to evaluate QC results

The substantial differences in CV, and correspondingly in SD, observed for different QC/reagent lot partitions

and among three identical specification instruments have implications for setting expected SD values used to evaluate individual QC results. QC results from short time intervals, for example, a few weeks or months, are not a good representation of “typical” performance. SDs estimated from short time interval data will likely cause inappropriate conclusions regarding acceptability of QC results using SD based evaluation rules. Consequently, conclusions may be incorrect regarding reporting or not-reporting results for PSs.

Our data suggest that best practice is to determine the SD as a pooled value from results partitioned by QC and reagent lot over a sufficient time interval such that these and other sources of variability are reflected in the estimate of SD used in QC rule evaluation. Note that pooling should be done with CVs (relative SDs) because the numeric values for a QC material vary with the reagent lot as well as the QC lot. The SD to use in QC rules is calculated from the pooled CV and target value for particular lots of QC and reagent in use. Consistent with recent recommendations [9, 10], a previously established long-term SD should be used with new lots of QC material. Attempting to re-establish a new estimate of SD for each lot of QC material is likely to cause an inappropriate estimate of SD and thus an inappropriate assessment of measuring system performance using SD based rules to interpret the QC results.

## Conclusions

The CV determined from QC data varied substantially over different QC and reagent lots for a single instrument and among different identical specification instruments. Estimating a CV or SD from QC data that represents the “typical” variability in a measuring system requires pooling data over long enough time intervals to include most sources of variability in the estimate. The SD for use in rules for QC result interpretation must be estimated from data over a long time interval with sufficient QC and reagent lot partitions so that a pooled SD value suitably reflects “typical” performance of a measuring system. Estimating the uncertainty of a laboratory result using the top-down approach based on results from internal QC data requires pooling QC data from a long time interval to ensure the uncertainty provided to users of the laboratory results represents “typical” measurement conditions. Caution is needed when using the estimated uncertainty of a result to make interpretations near decision values or based on reference change values because of the uncertainty in the estimated uncertainty of the result.

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

**Research funding:** None declared.

**Employment or leadership:** None declared.

**Honorarium:** None declared.

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

**Ethical approval:** Not applicable.

## References

1. Padoan A, Sciacovelli L, Zhou R, Plebani M. Extra-analytical sources of uncertainty: which ones really matter? *Clin Chem Lab Med* 2019;57:1488–93.
2. ISO 15189:2012. Medical laboratories — requirements for quality and competence. Geneva, Switzerland: International Organization for Standardization, 2012.
3. JCGM 100:2008. Evaluation of measurement data — guide to the expression of uncertainty in measurement. Sevres, France: Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, 2008.
4. Braga F, Panteghini M. The utility of measurement uncertainty in medical laboratories. *Clin Chem Lab Med* 2020;58:1407–13.
5. CLSI EP29. Expression of measurement uncertainty in laboratory medicine, 1st ed. Wayne, PA, USA: Clinical and Laboratory Institute, 2012.
6. ISO TS 20914:2019. Medical laboratories — practical guidance for the estimation of measurement uncertainty. Geneva, Switzerland: International Organization for Standardization, 2019.
7. Miller WG, Ereik 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.
8. Stavelin A, Riksheim BO, Christensen NG, Sandberg S. The importance of reagent lot registration in external quality assurance/proficiency testing schemes. *Clin Chem* 2016;62:708–15.
9. Miller WG, Sandberg S. Quality control of the analytical measurement process. In: Rifai N, Horvath AR, Wittwer C, editors. *Tietz textbook of clinical chemistry and molecular diagnostics*, 6th ed. Amsterdam, The Netherlands: Elsevier, 2017:121–56.
10. CLSI C24. Statistical quality control for quantitative measurement procedures: principles and definitions, 4th ed. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.
11. Plebani M, Padoan A, Sciacovelli L. Measurement uncertainty: light in the shadows. *Clin Chem Lab Med* 2020;58:1381–3.

---

**Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2020-0320>).