Marc H.M. Thelen*, Rob T.P. Jansen, Cas W. Weykamp, Herman Steigstra, Ron Meijer and Christa M. Cobbaert

Expressing analytical performance from multi-sample evaluation in laboratory EQA

DOI 10.1515/cclm-2016-0970
Received October 25, 2016; accepted December 20, 2016; previously published online February 9, 2017

Abstract

Background: To provide its participants with an external quality assessment system (EQAS) that can be used to check trueness, the Dutch EQAS organizer, Organization for Quality Assessment of Laboratory Diagnostics (SKML), has innovated its general chemistry scheme over the last decade by introducing fresh frozen commutable samples whose values were assigned by Joint Committee for Traceability in Laboratory Medicine (JCTLM)-listed reference laboratories using reference methods where possible. Here we present some important innovations in our feedback reports that allow participants to judge whether their trueness and imprecision meet predefined analytical performance specifications.

Methods: Sigma metrics are used to calculate performance indicators named 'sigma values'. Tolerance intervals are based on both Total Error allowable (TEa) according to biological variation data and state of the art (SA) in line with the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Milan consensus.

Results: The existing SKML feedback reports that express trueness as the agreement between the regression line through the results of the last 12 months and the values obtained from reference laboratories and calculate imprecision from the residuals of the regression line are now enriched with sigma values calculated from the degree to which the combination of trueness and imprecision are within tolerance limits. The information and its conclusion to a simple two-point scoring system are also graphically represented in addition to the existing difference plot.

Conclusions: By adding sigma metrics-based performance evaluation in relation to both TEa and SA tolerance intervals to its EQAS schemes, SKML provides its participants with a powerful and actionable check on accuracy.

Keywords: analytical performance specifications; bias; external quality assessment; imprecision.

***Corresponding author: Marc H.M. Thelen,** Laboratory for Clinical Chemistry, Amphia Hospital, Postbox 90158, 4800RK, Breda, The Netherlands, Phone: +31765952030, Fax: +31765953807, E-mail: mthelen@amphia.nl; and SKML, Organization for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands

Rob T.P. Jansen, Herman Steigstra and Ron Meijer: SKML, Organization for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands

Cas W. Weykamp: SKML, Organization for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands; and Laboratory for Clinical Chemistry, Queen Beatrix Hospital, Winterswijk, The Netherlands

Christa M. Cobbaert: SKML, Organization for Quality Assurance of Medical Laboratory Diagnostics, Radboud University, Nijmegen, The Netherlands; and Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, The Netherlands

Introduction

Independent verification of metrological traceability of in vitro medical diagnostic (IVD) tests and predefined tolerance intervals for errors of measurement to evaluate whether medical tests are fit-for-purpose, are recognized as the fifth and sixth, respectively pillar of the temple of laboratory standardization, beyond the establishment of Reference Materials, Reference Methods, accredited Reference Laboratories and traceable Reference Intervals and Decision Limits [1]. Over the last years, the Dutch EQA organizer, Organization for Quality Assessment of Laboratory Diagnostics (SKML), has innovated its external quality assessment scheme (EQAS) for general clinical chemistry towards a category 1 EQAS scheme [2, 3] with commutable samples with value assignment in reference laboratories. Here we present the added value of "multi-sample evalution (MUSE)", a new reporting and scoring system that has been developed as a quality performance tool of the accuracy-based EQA scheme that supports corrective actions of the participant and allows the evaluation of the success of previous corrective actions.

Most medical laboratories have built their quality system on International Organization for Standardization (ISO15189:2012), which requires participation in an EQAS. Analytical performance of EQA samples is judged against predefined quality criteria for bias and imprecision,

derived from either (outcome-defined) medical decision criteria, biological variation, or from state-of-the-art performance [4]. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) has suggested 25 performance indicators to monitor and benchmark laboratory quality; 2 of the 25 indicators are comprised the results of EQA [5].

SKML has chosen to use commutable EQA samples value assigned by accredited reference laboratories using Joint Committee for Traceability in Laboratory Medicine (JCTLM)-listed reference methods and/or materials [6, 7] when available. For participants in the general chemistry EQAS scheme, this has resulted in smaller between-laboratory imprecision and better agreement with reference methods [3, 8, 9]. Although some EQAS schemes ask for analysis of single samples in duplicate and report the imprecision of these duplicates, the lack of coherence between individual samples still impedes quantification of the bias component of the inaccuracy. Therefore, SKML has developed a reporting system that is based on moving regression analysis of multiple samples. What the existing SKML reporting system still lacked was a way to quantify laboratory performance against predefined specifications. Here we present 'MUSE' as a further development of our reporting system, now adding the following aspects to the existing reports: (1) sigma metrics using a total error allowable (TEa) sigma value with a tolerance interval derived from biological variation. (2) Sigma metrics using an SA sigma value based on a state-of-the-art tolerance interval. (3) Graphical representation of both sigma values in the difference plot that displays the regression of the participants findings with the target values in a background of TEa and state-of-the-art (SA) tolerance intervals. (4) A simple two point scoring system that depicts at a glance whether the sigma value that is applicable for a certain measurand is either appropriate (green) or reason for corrective action (red).

Materials and methods

Schemes and samples

We show examples of our scheme with most participants (approximately 125 participants with approximately 250 participating instruments in total), the general chemistry scheme, in which participants receive each year 24 samples consisting of blind duplicates of 12 individual samples with different concentrations for all general chemistry components. Samples spanning the entire clinically relevant concentration range are used. Participants are requested to analyze one sample every 2 weeks.

As published before [3, 6, 8], SKML samples are commutable and their values were assigned using JCTLM-listed reference systems where possible. If there are no reference values available, target values – determined by expert laboratories – or consensus method group averages are used.

Reports

All reports show both evaluation of only those results that are new relative to previous reports and long-term evaluation concerning all results of the last 12 months. In case of the general chemistry scheme, results are reported after analysis of six samples (3 months) with the long-term evaluation concerning 24 samples.

The most important features of the SKML reporting system MUSE are listed below.

Regression analysis

Regression lines of laboratory results against target values (reference values, expert laboratory values or consensus method group averages) are time-weighted, with the most recent results receiving the greatest weight in the calculation of the regression line. Two regression lines are calculated: one for the results of all samples within the last 12 months and one for the last reporting period. For the general chemistry, this results in regression lines and statistics for both the last year (24 samples) and the 3 months (6 samples).

Tolerance ranges

Scores are assigned on the basis of two tolerance ranges, in which results must be located: the TEa tolerance range and the SA tolerance range.

The SA tolerance range is a function of the concentration, with a shape determined by the analytical precision profile (see Supplemental Data) and is determined every 3 years. The SA tolerance range has a width of 3 SDsa. SDsa is the state-of-the-art SD as calculated from all participants' results after clean-up of results that cannot be considered 'state of the art' after outlier removal. For details on the outlier removal procedure, we refer to the online Supplemental Data. The clean-up process that renders the data of 'all users' to 'state of the art' is performed by experts of the individual SKML schemes. The criterion that decides whether results of a particular method are excluded from the SA calculation is whether more specific methods are available. In the general chemistry scheme this has resulted in exclusion of all Jaffé based methods for creatinine from the SA precision profile.

According to the Stockholm [10] and later Milan criteria [4], the TEa tolerance range is based on the desirable specifications database based on biological variation as published by Ricos et al. [11]. The TEa value from the database is used at the so-called target level. The target level is determined as the heart of the typical reference interval for each measurand. Since no evidence-based models are available for the extrapolation of the TEa to other concentration levels we have chosen to extrapolate that value to other concentration levels by using the same shape of the profile as calculated from the SDsa profile. (see Supplemental Data). Therefore, the TEa tolerance profiles

have the same shape as the SDsa tolerance profiles, but a different width, which is determined by biological variation.

Outlier removal

Two different forms of outlier removal are applied. The first approach excludes results that are unlikely to belong to a method group of results of different laboratories for that same sample. The second approach excludes results based on analysis of results of a single participant after comparing the correlation with other results in same challenge. For details we refer to the online Supplemental Data.

Time weighting

The most recent values receive more weight in the calculations than results further back in time. The weighting parameter W_i is calculated as:

$$W_i = 2^{-\Delta t/\alpha}$$

Δt is the time (expressed in months) between the submission deadline of the last survey and the measurement date of sample i. The factor α is the half-life (also expressed in months) and is by default 6 months, resulting in a weight for a result from a year before of 25% compared to the latest.

Within-laboratory SD

The within-laboratory SD is calculated as the residual SD of the time-weighted regression line through the laboratory results versus the target values.

Between-laboratory SD

For every sample, the between-laboratory standard deviation (SDbl) is calculated from the total standard deviation (SDt) and the average within-laboratory standard deviation (SDwl) at the concentration level of the sample. For this purpose, the average within-laboratory SDwl is extrapolated to the concentration level of the sample using the precision profile.

Sigma values

The MUSE scoring system uses the sigma metrics concept. This is used world-wide to quantify the quality of a production process. In a process that meets the requirements of the Six-Sigma standard, the scatter is so low that less than one in million products do not meet the quality standard. This means that for a six sigma process 6SD's are within the tolerance limits set. The Six-Sigma concept accepts a shift of 1.5 sigma after some time. For this reason the tolerance limits used are based on 4.5 sigma rather than on six sigma. Sigma values are calculated for both the TEa and the SDsa tolerance limits and both are calculated for the cumulative long-term regression as well as for the short term regression. For details see the online Supplemental Data.

Scores based on sigma values: To indicate whether a sigma value is acceptable or not, we have introduced a scoring system. Participants earn two points for a TEa sigma value of at least 4.5. A TEa sigma values between 2 and 4, 5 gives rise to 1 point. Lower sigma values result in zero points which are marked in red background color, whereas 1 and 2 points are presented on a green background. When SDsa tolerance limits are wider than TEa tolerance limits, scores are based on SDsa sigma values to prevent participants from being frustrated by scores which they cannot improve with currently available methods.

For details on the calculations of within- and between-laboratory statistics including the sigma metrics, we refer to the online Supplemental Data.

Results

Implementation of multi-sample evaluation of trueness and precision in MUSE has resulted in a new system for graphical representation of results with a new mathematical system for representing moving trueness and precision information and a scoring system based on the multi-sample approach. We describe and discuss each individual element of this reporting and scoring system, respectively.

The difference plot

In the difference plot the participants find their results of the last reporting period as yellow squares with characters in alphabetical order representing the individual samples evaluated in this reporting period. Figure 1 shows an example from the general chemistry scheme which is reported after analysis of every six samples. The first analyzed samples of the period is marked as 'A', and the most recently analyzed sample is marked as 'F'. Results of previous samples are represented by smaller yellow squares without characters. The results are plotted in a difference plot with the value assigned by the reference laboratory on the x-axis and the deviation of the participant results in absolute units on the y-axis. As a representation of the moving trueness, a time-weighted line is fitted through all yellow points of the last year, with newer points having more weight (for details see Supplemental Data). By providing a moving reflection of all samples of the last 12 months, we have added the reflective value of an annual report to every single report. To allow for evaluation of earlier corrective actions, an additional time weighted line is fitted through the letter-marked newest yellow points of the current reporting period. The rationale for complicating the regression by adding time weighting was made to combine the advantage of adding historical points for

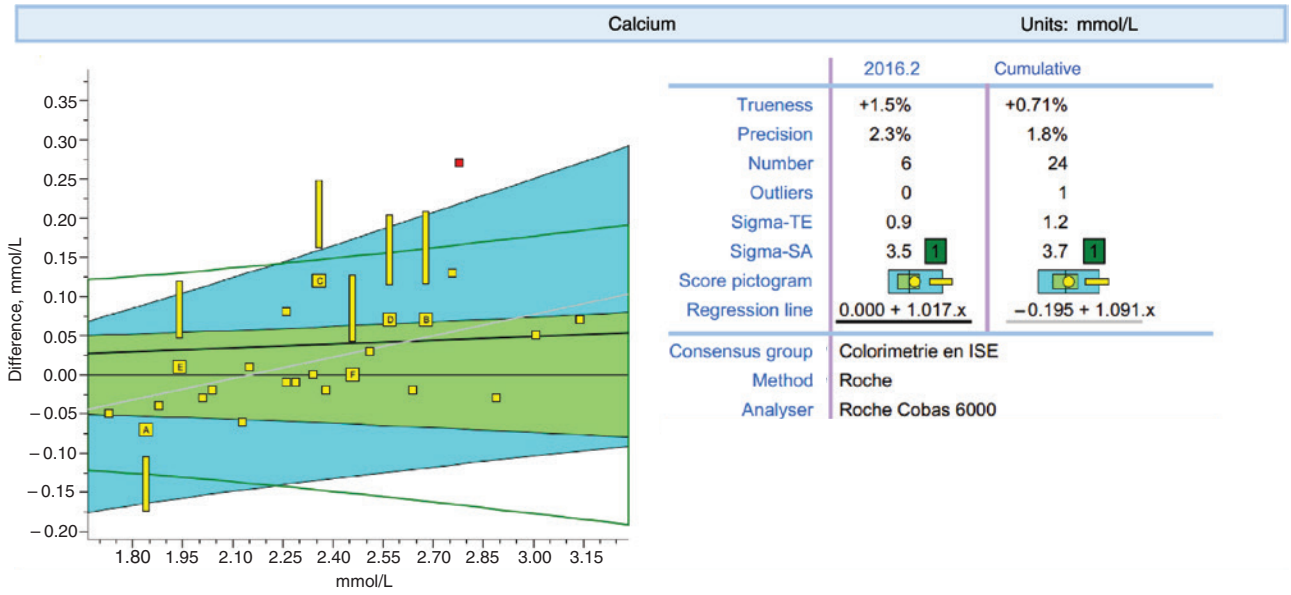


Figure 1: Difference plot and scoring table.
A typical difference plot for calcium of a participating laboratory. The green lines represent the tolerance interval based on SDsa imprecision and at reference value trueness. The red point indicates an outlier in a previous challenge. Next to the difference plot mathematical information on both regression lines is presented as, bias, imprecision, TE sigma score, SA sigma score, number of contributing samples and linear regression line formula compared to reference. The resulting score of one point is color coded in green next to the SA sigma value on which it is based for this measurand. In the score pictogram the green and blue boxes symbolize the TEa tolerance range and the SA tolerance range, respectively. The location of the average value of this participant is represented by the yellow circle and is therefore a representation of the eccentricity. The yellow bar is the sigma scale (2.0–4.5), which can be read from the edges of the TEa and SA tolerance ranges.

better statistics with the emphasis of the most recent data in the long-term regression line.

Note that in the example shown, the blue SA tolerance area is shifted relative to the TEa area that is anchored on the reference values, due to bias in the state-of-the-art results. The presented laboratory has apparently chosen to correct its bias. The success of this intervention can be judged by that participant by comparing the short term black regression line, with that of the long-term gray line that is partly determined by older points that were measured before the bias correction.

The relative distances of the individual yellow points from the regression line represent the imprecision of the participating laboratory. Since every year, 24 samples of our general chemistry scheme consist of 12 sets of blinded duplicates, we were able to compare imprecision calculated from dispersion around the regression line (residuals) to imprecision as calculated from the coefficient of variation of the 12 duplicates. Results were all within the mutual confidence intervals. (Data not shown).

To depict the imprecision of the individual points, we have chosen to plot yellow bars that represent the area beginning at two sigma (closest to the regression line) and ending at 4.5 sigma value (most remote from the regression line) of the particular point. The imprecision can be judged on the blue and green background of the graph. The sigma score of each single point is calculated as that value where the yellow bar crosses the tolerance area, resulting in a sigma TEa at the cross-point with the green area and a sigma SDsa at the cross point with the blue area. Better precision results in shorter yellow bars, reflecting higher sigma values to fit within the tolerance area. The sigma score of an evaluation period is calculated as the mean of all points involved after time weighting (see online Supplemental Data for details on time weighting).

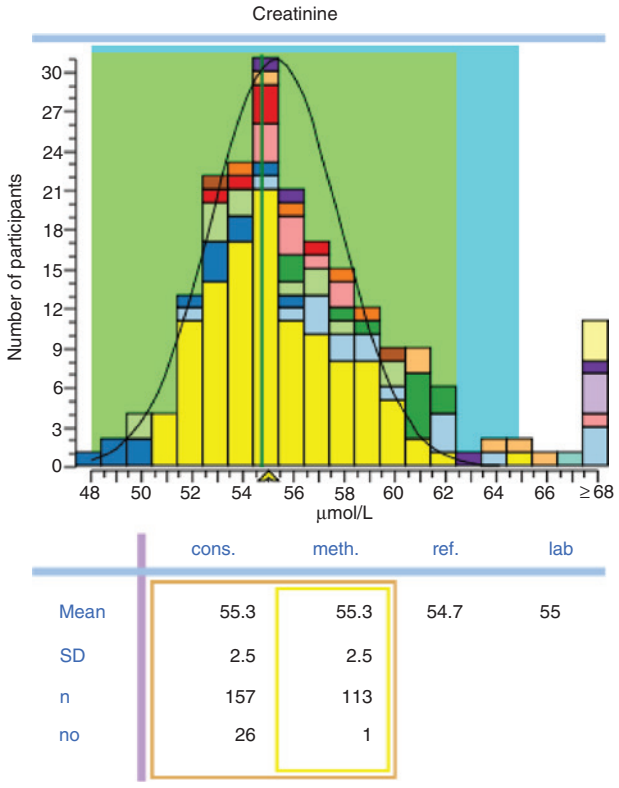
The blue area represents the tolerance interval based on the state of the art precision profile. The green area represents the tolerance interval based on TEa. Scores are only set on the basis of state-of-the-art in cases where currently available methods do not allow to achieve goals based on TEa criteria. If scores are given on the basis of state-of-the-art rather than TEa, two green lines are shown indicating a widened TE area to SA width, combining state-of-the-art imprecision tolerance with bias tolerance based on reference values. An example of these added green lines is shown in Figure 1.

Next to the difference plot, numerical values of the information presented in the difference graph are displayed

in a table (Figure 1). Also in the table is a score that can be 1 or 2 points, which are also depicted in the score indicator. Whether the score is based on the blue state-of-the-art tolerance limit or on the green TEa limit is decided by the question whether the state of the art is wider than the TEa area or not. All information of difference plot and score of the current reporting period is depicted in a condensed format in the score pictogram in red or green.

Histograms

In the histogram (Figure 2) section, the participants find the result of every single sample of the last survey plotted as a yellow triangle under the value x-axis. The histograms plot the frequency of the results of all methods on the y-axis and show the Gaussian curve fitted through the results of the method group. This will help the participant understanding the cause of deviation from the reference.



	cons.	meth.	ref.	lab
Mean	55.3	55.3	54.7	55
SD	2.5	2.5		
n	157	113		
no	26	1		

Figure 2: Method histogram.
For each sample a histogram shows the distribution of results between methods. A Gaussian curve is fitted through the data of each users methods in the histogram of one sample of a reporting period, in this case for Creatinine. Results lying at ≥ 3 SD after finding the best fit are considered outliers. Yellow arrow on x-axis: your result; yellow bars: your method; other colors represent other combinations of manufacturer and method. Under each histogram statistical information is provided that compares the value found in the reference laboratory (ref) to those of the participant (lab), the average of its method group (meth) and that of all users (cons). For each sample the mean results, its standard deviation (SD), the number of results involved (n) and the number of outliers (no) is reported.

Survey and score summary

As a cover page (Figure 3) of every survey report, MUSE depicts an executive summary of the findings in the current reporting period compared to the reference. It also shows graphics of score pictograms and score indicators for current and historical reports. This facilitates both swift discovery of measurands that need attention and review of effectiveness of corrective actions, essential for a vivid plan-do-check-act cycle for continuous improvement. From this summary, determining which measurands are scored on the TE sigma values and which on the SA sigma values is possible. When the green TEa tolerance range is wider than the blue SA tolerance range, scores are based on TE sigma values, in other cases on SA sigma values. As can be seen in the case of the presented general chemistry scheme, 12 measurands are scored on each of the tolerance limits. Also, it can be seen that for six of the 24 measurands no reference method is available and trueness is therefore judged against consensus. In these cases, paler shades of blue and green are used in the score summary and the corresponding difference plot for the particular measurands.

Discussion

To check whether the IVD industry has successfully implemented the metrological traceability concept which is needed for standardization of medical tests, EQAS organizers are needed to verify the accuracy of metrological traceability. Therefore, SKML developed an accuracy-based EQA scheme for general clinical chemistry measurands and an advanced MUSE reporting and scoring system that gives measure to analytical performance with insight in the individual contribution of bias and imprecision to total error.

By separating imprecision from bias, MUSE allows laboratories to recognize the constant part of bias. This makes EQAS participation actionable, as ISO15189 writers had in mind when they demand EQAS participation as a tool for quality improvement of medical diagnostic laboratories. The fact that every single MUSE report also

Analyte		Trueness				Precision		Performance				
		your mean	ref.	cons.	SDbl	your prec.	SDwl	this survey	PS	cumulative	PSc	
Urea	mmol/L	15.8		15.5	0.5	0.6	0.4		2			2
Creatinine	µmol/L	155.1	152.4	151.5	4.3	3.1	3.0		1			1
Sodium	mmol/L	140.2	143.8	142.7	1.5	1.8	1.2		1			1
Potassium	mmol/L	5.21	5.37	5.39	0.07	0.09	0.06		1			1
Chloride	mmol/L	100.9	102.3	100.8	1.8	1.2	1.0		2			2
Calcium	mmol/L	2.377	2.230	2.242	0.060	0.117	0.042		0			1
Inorg. Phosphate	mmol/L	1.635		1.574	0.034	0.066	0.029		1			1
Magnesium	mmol/L	1.204	1.167	1.163	0.029	0.076	0.024		1			1
Urate	mmol/L	0.336	0.348	0.346	0.010	0.015	0.007		2			2
Bilirubin	µmol/L	40.1	37.1	36.6	1.3	2.4	1.1		1			2
ASAT	U/L	72.1	74.6	74.6	2.9	2.5	1.7		2			2
ALAT	U/L	76.3	82.3	80.6	3.8	2.7	2.0		2			2
LD	U/L	390	448	454	30	27	16		0			0
Alk. Phosphatase	U/L	144	155	152	9	8	5		2			2
Gamma-GT	U/L	68.9	67.5	68.0	2.4	2.0	1.7		2			2
Amylase	U/L	186	189	184	7	8	4		2			2
Lipase	U/L	25.2		26.8	1.3	1.4	0.9		2			2
CK	U/L	231	245	243	11	6	5		2			2
Total Protein	g/L	65.9	65.4	65.5	1.4	1.6	1.0		2			1
eGFR (F, 55, white)	mL/min/1.73m²	23.6	24.1	23.0	1.2	0.9	0.7		2			2
Glucose	mmol/L	13.85	13.39	13.56	0.35	1.15	0.28		1			1
Albumin	g/L	46.1		46.7	1.3	1.6	0.9		1			1
Osmolality	mOsmol/kg	310.6		313.9	3.9	5.4	2.9		1			1
Iron	µmol/L	36.2		36.1	0.9	0.9	0.8		2			2
								Total :	35			36

Figure 3: Summary sheet.
The summary sheet presents a quick overview with mean findings in the current reporting period for every measurand against reference findings and score pictograms and score indicators for every measurand in survey. Display of historical scores allows for review of success of corrective action. The summary sheet shows, respectively; under trueness: the mean of the laboratory findings (your mean), the mean as found in the reference laboratory (ref), the mean of the consensus group (cons), and the between-laboratory precision (SDbl). Under precision: the precision of the laboratory (your prec), the mean within-laboratory precision of all participants (SDwl). Under performance: the participant's performance score (SC) of this survey and the participant's cumulative performance score (PSc) over all samples of the last year.

contains incremental and moving information of the samples of the last 12 months enriches these reports with information that other EQA schemes reserve for annual reports. We realize that guaranteeing traceability to reference methods is a responsibility of the IVD providers according to both FDA and CE marking of IVDs. Laobratories, however, have a ISO15189 responsibility to verify the metrological traceability as defined by ISO17511. Between the report on a possible traceability issue by a laboratory and the corrective action by the IVD provider, laboratories need actionable information as reported in MUSE that can guide their corrective actions. When histograms show that differences are caused by a problem in a method SKML contacts, the IVD provider involved and advises participants on method based bias correction awaiting global action from the IVD manufacturer.

ISO17043 is the international standard for accreditation of EQAS organizers and it demands a poor performer policy. A performer policy requires a scoring system with underlying performance evaluation and accompanying predefined tolerance limits with a clear rationale needed. To allow for undisputed authority EQAS samples must be commutable and value assigned in JCTLM-listed reference laboratories. In the 2014 Milan conference, performance goals based on biological variation are set as the standard for all cases where criteria based on clinical outcome are unavailable. We have built MUSE in line with that thought. In cases where criteria on clinical outcome are available, we have based our TEa values on those criteria. Until now, this is only implemented for cardiac troponin-T. In all other cases TEa is based on biological variation data. Since biological variation has to be taken into account when calculating whether a result differs significantly from a previous result by calculating the reference change value [12], biological variation also seems to play a role in judging the significance of the clinical course.

The Milan criteria for performance goals also mention state-of-the-art criteria as an alternative. In line with that, we use state-of-the-art criteria next to biological variation criteria for those measurands where no methods are

currently available that meet the performance goals set by biological variation. In those cases, we base the participant score on the SA sigma, but we still report the TE-sigma along with the SA sigma in order to help participants and scientific societies in their dialogue with IVD-manufacturers for better assays. This makes the scores fair to the users, which is important for acceptance.

In the total error concept the TE budget can be spent on either bias or imprecision [13, 14]. Discussion [14] on the question whether current mathematical models should be revised does not interfere with the concept that EQAS results can be judged on a concept that allows for an error budget that can be spent on either bias or imprecision or both. SKML participants of a scheme using value-assigned commutable materials may strive for (temporary) recalibration of their assays towards the levels founds in the reference methods. This will leave the participants with their complete total error budget to be spent on imprecision [15].

Since many participants also share their internal QC data with us, we can study the agreement between imprecision in internal QC and imprecision as calculated from the external QC regression residuals. For all measurands, the within-laboratory imprecision as calculated from internal controls was similar, but not identical to imprecision as calculated from the external controls as explained. Differences may be partly explained by difference in concentration levels, partly by the non-commutability of the internal controls, and partly by the more robust statistics on the larger number of data for internal controls. However, typically laboratories with good precision for internal controls, also show good precision for external controls.

Like every reporting system, the approach of MUSE has limitations. Since there are no guidelines on how to calculate bias from multi-sample regression analysis and how to calculate imprecision from the residuals the choices that are made by SKML in their mathematical approach may be classified as arbitrary. However, the experience that SKML creates with MUSE presents an opportunity for evaluation of these choices, allowing incremental improvement of such considerations. Another limitation is that the MUSE report lacks information on the uncertainty of the estimate of bias and imprecision as a result of lack of uncertainty of the regression as determined by a combination of the uncertainty of the value assignment and the uncertainty inherent to the mathematical approach. We have judged that the added value of such information is not enough to justify the added complexity it brings to the reports at this time. Another limitation is the arbitrary choice to extrapolate the TEa tolerance levels based on biological variation from the mean of the reference interval to other concentration levels with a shape similar to that of the SA precision profile. When better models for this extrapolation become available for the TEa tolerance profiles we could adopt those. Customer satisfaction evaluation over time has to learn whether participants take enough advantage of the performance quantification to accept the necessary effort to fully understand and appreciate the reports. Spontaneous customer feedback so far teaches us that both are true; the intentions are appreciated, but they do require effort to understand. Positive feedback concentrates on the score pictograms and summary sheets that allow for a quick overview and guidance to the areas that need attention. The feedback on the sigma values is mixed. On the positive side, participants appreciate that the tolerance intervals have a rationale; on the down-side most participants would like to be able to fully understand the mathematics and even would like to be able to calculate the sigma values by themselves which seems unachievable given the time weighting in the regression.

In summary, we hope that SKML MUSE will inspire other EQAS organizers to apply a multi-sample evaluation approach to their current schemes.

Schemes with commutable native sera with value assignment by reference methods allow for verification of metrological traceability of IVD tests and predefined tolerance limits of measurement errors to evaluate whether medical tests are fit-for-purpose. We conclude that our reporting and scoring system provides an important contribution to the missing pillars for completing the temple of laboratory standardization [1].

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. Braga F, Panteghini M. Verification of in vitro medical diagnostics (IVD) metrological traceability: responsibilities and strategies. Clin Chim Acta 2014;432:55–61.
2. Miller GW, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670–80.

3. Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, et al. A category 1 EQA scheme for comparison of laboratory performance and method performance: an international pilot study in the framework of the Calibration 2000 project. Clin Chim Acta 2014;432:90–8.
4. Sandberg S, Fraser C, Horvath A, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–5.
5. Sciacovelli L, O'Kane M, Skaik YA, Caciagli P, Pellegrini C, Da Rin G, et al. Quality Indicators in Laboratory Medicine: From theory to practice: Preliminary data from the IFCC Working Group Project "laboratory Errors and Patient Safety". Clin Chem Lab Med 2011;49:835–44.
6. Weykamp C, Franck P, Klein Gunnewiek J, de Jonge R, Kuypers A, van Loon D, et al. Harmonisation of seven common enzyme results through EQA. Clin Chem Lab Med 2014;52:1549–55.
7. Drion I, Cobbaert C, Groenier KH, Weykamp C, Bilo H, Wetzels J, et al. Clinical evaluation of analytical variations in serum creatinine measurements: why laboratories should abandon Jaffe techniques. BMC Nephrol 2012;13:133.
8. Cobbaert C, Weykamp C, Franck P, de Jonge R, Kuypers A, Steigstra H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples-Five year experience from the Netherlands. Clin Chim Acta 2012;414:234–40.
9. Perich C, Ricós C, Alvarez V, Biosca C, Boned B, Cava F, et al. External quality assurance programs as a tool for verifying standardization of measurement procedures: pilot collaboration in Europe. Clin Chim Acta 2014;432:82–9.
10. Fraser C. General strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999;59:487–90.
11. Minchinela J, Ricós C, Perich C, Fernández-Calle P, Alvarez V, Domenech M, et al. Biological variation database, and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update http://www.westgard.com/biodatabase1.htm biodatabase-2014-update. Accessed 14 Feb 2016.
12. Fraser C. Reference change values. Clin Chem Lab Med 2011;50:807–12.
13. Petersen P, Jørgensen L, Brandslund I, De Fine Olivarius N, Stahl M. Consequences of bias and imprecision in measurements of glucose and Hba1c for the diagnosis and prognosis of diabetes mellitus. Scand J Clin Lab Invest Suppl 2005;240:51–60.
14. Oosterhuis W. Gross overestimation of total allowable error based on biological variation. Clin Chem 2011;57:1334–6.
15. Westgard J. Advanced quality management/six SIGMA Translating Method Validation Data into Sigma Metrics. https://www.westgard.com/lesson78.htm. Accessed 14 Feb 2016.

Supplemental Material: The online version of this article (DOI: 10.1515/cclm-2016-0970) offers supplementary material, available to authorized users.