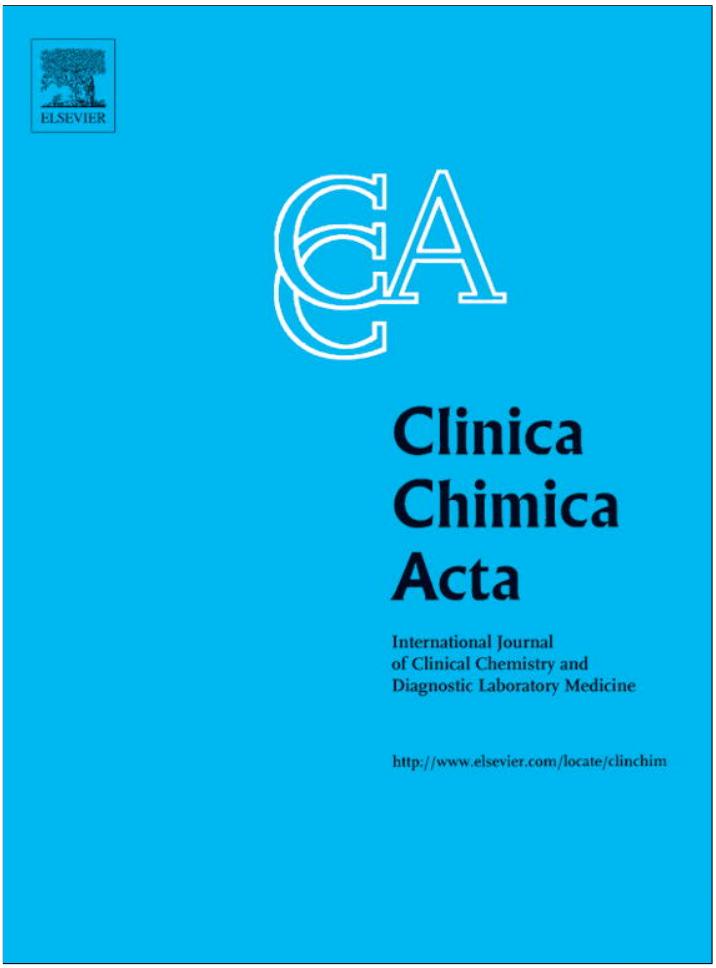

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Clinica Chimica Acta 414 (2012) 234–240

Contents lists available at SciVerse ScienceDirect

Clinica Chimica Acta

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

Systematic monitoring of standardization and harmonization status with commutable EQA-samples—Five year experience from the Netherlands

Christa Cobbaert [a,*], Cas Weykamp [b], Paul Franck [c], Robert de Jonge [d], Aldy Kuypers [e], Herman Steigstra [f], Jacqueline Klein Gunnewiek [g], Douwe van Loon [h], Rob Jansen [f]

[a] Leiden University Medical Center, Leiden, The Netherlands
[b] Beatrix Hospital, Winterswijk, The Netherlands
[c] Haga Hospital, Den Haag, The Netherlands
[d] EMCR, Rotterdam, The Netherlands
[e] Maasziekenhuis Pantein, Boxmeer, The Netherlands
[f] SKML, Nijmegen, The Netherlands
[g] Ziekenhuis Gelderse Vallei, Ede, The Netherlands
[h] Sint Antoniusziekenhuis, Nieuwegein, The Netherlands

ARTICLE INFO

Article history:
Received 5 August 2012
Received in revised form 26 August 2012
Accepted 24 September 2012
Available online 2 October 2012

Keywords:
Standardization
Harmonization
Commutable EQA-materials
Desirable bias and imprecision
Total allowable error
Test equivalence

ABSTRACT

Background: Equivalence of results among laboratories is a major mission for medical laboratories. Monitoring of test equivalence is structurally integrated in the Dutch External Quality Assessment (EQA) scheme since 2005. Commutable poolsera, single donation "spy" sera and biological variance tolerance limits have been introduced in the EQA scheme for evaluation of the degree of test equivalence and its determinants.
Methods: In the annual cycle scheme 24 samples, covering the (patho)physiological measuring range for 17 analytes, are assayed by 220 participating laboratories at biweekly intervals. Test equivalence was evaluated by calculating overall median interlaboratory coefficients of variation (CVs) and its bias and imprecision components. Data from 2005 and 2010 schemes are evaluated to investigate trends in performance and success of standardization efforts.
Results: Overall median interlaboratory CVs in 2010 were mostly better than in 2005. Median interlaboratory CVs became <5% for electrolytes and substrates, and <10% for enzymes. Improvement in median interlaboratory CVs over these five years is mainly explained by improved method standardization, especially for enzymes and creatinine.
Conclusion: The Dutch EQA-program proves to be a powerful instrument to evaluate test equivalence. It allows monitoring standardization efforts in a highly effective way and gives insight into remaining standardization potential.

1. Introduction

Interchangeability of laboratory test results across laboratories and time is a major topic in laboratory medicine and can be achieved by either standardization or harmonization [1–4]. The degree of interchangeability or test equivalence and the success of standardization/harmonization efforts can be monitored by external quality assessment (EQA) schemes, also known as proficiency testing programs [4]. Major advantages of using EQA schemes are that these a) reflect real life analytical conditions as ideal research circumstances are avoided, b) provide robust data as many labs and many methods are included and c) can be organized efficiently without requiring separate evaluations for monitoring harmonization/standardization. To be an effective monitoring tool for assessing traceability, EQA schemes should meet at least two fundamental requirements. Firstly, the EQA-specimens used should be commutable—i.e. behave like native patient materials—to prevent that differences seen are related to matrix effects rather than to differences between methods. Secondly, the target value should, whenever feasible, preferentially be assigned by JCTLM-listed reference laboratories with approved reference systems. Value assignment can be done either directly with a reference measurement procedure or a designated comparison method, or indirectly by anchoring the assigned value to a certified reference material under the condition that transferability is guaranteed. In addition biological variance based tolerance limits should be used. According to the Stockholm consensus conference on quality specifications in

Abbreviations: CV, coefficient of variation; EQA, external quality assessment; IVD, in vitro diagnostic; JCTLM, Joint Committee on Traceability in Laboratory Medicine; SKML, the Dutch EQA, named Stichting Kwaliteitsbewaking Medische Laboratorium Diagnostiek; TE_a, allowable total error.

* Corresponding author at: Leiden University Medical Center, Department of Clinical Chemistry, E2-28, Albinusdreef 2, 2300 RC Leiden, The Netherlands. Tel.: +31 71 526 4483.
E-mail address: c.m.cobbaert@lumc.nl (C. Cobbaert).

0009-8981/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cca.2012.09.027

laboratory medicine [5], EQA consensus results are on the 5th level of its hierarchy and biological variation based criteria on its 2nd level. EQA schemes in other countries have set Minimal Allowable Performance limits, sometimes based on consensus group mean values and tolerance limits based on e.g. 95th percentile of results [6], other based on biological variation [7]. However, no information on the commutability of the samples used is given.

SKML (Stichting Kwaliteitsbewaking Medische Laboratorium Diagnostiek), the EQA provider in the Netherlands, organizes EQA schemes meeting these requirements since 2005 [8–18]. In addition, SKML integrates standardization and harmonization efforts since 1998 under the flagship of Calibration 2000 for analytes with unacceptable bias in the EQA scheme [11–15,18]. Thirdly, a scoring system was developed based on biological variation. In this paper national general clinical chemistry data of the EQA schemes in 2005 and 2010 are compared for 17 parameters to investigate a) whether analytical performances have improved, b) whether standardization efforts have been successful and c) whether there is room for further improvement of equivalence. For medical lab professionals these data are an appropriate means to verify if the in vitro diagnostic (IVD)-industry meets the IVD directive 98/79/EC. This European directive obliges manufacturers to produce kits with traceable measurement results and documented uncertainty. The aggregated data in this paper allow evaluating if the present IVD-kits indeed meet the medical needs. And when not, whether better quality can be achieved by more strict standardization or that intrinsically better methods are required to achieve quality goals.

2. Materials and methods

2.1. Specimens

Samples (N = 24) are prepared from fresh, anonymized left-over sera of the routine clinical chemistry laboratory with exclusion of icteric and lipemic samples. Left-over sera are tested for HBsAg, a-HIV and a-HCV and negative sera are stored frozen at −84 °C in aliquots of 200 mL. The use of anonymous left-over sera is in accordance with national guidelines on acceptable use of body fluids, and does not demand informed patient consent.

Prior to manufacture of the EQA samples the aliquots are thawed and pooled. Physiological and pathophysiological concentration ranges are created by adequately mixing pools and by spiking with minerals, recombinant human enzymes and human albumin. The concentration ranges that are systematically tested are presented in Table 1. After dispensing, vials are frozen at −84 °C until shipment to the participants. At the beginning of the annual cycle samples are shipped on dry ice to the participants who store them at −84 °C until analysis. Commutability of the Dutch EQA-samples has been established [16–19], and reference [14] with proof of commutability for 17 analytes, has been translated and summarized in an attached supplemental file. Throughout the years commutability has been monitored by including a native, single donation spy-sample that is prepared according to NCCLS C37-A2.

2.2. Target value assignment

Target values are set by JCTLM-endorsed Reference Laboratories using approved reference measurement procedures (www.bipm.org) in 13 out of 17 general chemistry analytes. Value assignments are systematically done in the low and the high pools for 13 constituents. The in-between levels are manufactured by mixing high and low pools in different amounts. The latter procedure allows calculating the target values for the in-between levels. Table 1 lists the respective general clinical chemistry analytes, the reference or definitive measurement procedures and the involved reference laboratories.

2.3. EQA-design

Since 2005 the Dutch EQA-scheme has used an EQA-toolbox, consisting of commutable, value-assigned EQA-materials and a scoring system based on biological variation, for monitoring metrological traceability.

The EQA scheme is framed in an annual cycle with 12 blinded samples measured for 17 parameters at two-weekly intervals in the first half year, and 12 blinded duplicate samples measured at two-weekly intervals in the second half year. By covering the physiological and pathophysiological concentration range twice for each parameter, the design allows to investigate duplicability, linearity and recovery.

Table 1
Clinical chemistry parameters tested in the Dutch EQA on analytical performance trends between 2005 and 2010.

Analyte		Symbol	Concentration range	Reference methods	Reference laboratory
Minerals	Calcium	Ca^{2+}	1.77–3.27 mmol/L	Atomic absorption spectrometry	INSTAND e.V., Düsseldorf, Germany
	Chloride	Cl^{a}	83–116 mmol/L	Coulometry	INSTAND e.V., Düsseldorf, Germany
	Magnesium	Mg^{2+}	0.59–2.01 mmol/L	Atomic absorption spectrometry	INSTAND e.V., Düsseldorf, Germany
	Potassium	K^{+}	3.2–7.8 mmol/L	Flame emission spectrometry	INSTAND e.V., Düsseldorf, Germany
	Sodium	Na^{+}	118 − 167 mmol/L	Flame emission spectrometry	INSTAND e.V., Düsseldorf, Germany
Substrates	Creatinine	Crea	54–262 μmol/L	GC-IDMS	DGKL, Bonn, Germany
	Glucose	Glu	3.9–30.0 mmol/L	GC-IDMS	INSTAND e.V., Düsseldorf, Germany
	Total Protein	TE	49–82 g/L	Modified Biuret Method	INSTAND e.V., Düsseldorf, Germany
	Uric Acid	UA	0.22–0.58 mmol/L	HPLC	Erasmus Medical Centre, Rotterdam, Netherlands
Enzymes	ALT	ALT	17–214 U/L (at 37 °C)	IFCC primary reference method; Clin Chem Lab Med 2002;40:718-24	Haga Hospital, The Hague, The Netherlands
	AST	AST	18–147 U/L (at 37 °C)	IFCC primary reference method; Clin Chem Lab Med 2002;40:725-33	Haga Hospital, The Hague, The Netherlands
	γ-GT	GGT	30–175 U/L (at 37 °C)	IFCC primary reference method; Clin Chem Lab Med 2002;40:734-38	Haga Hospital, The Hague, The Netherlands
	LDH	LDH	116–1143 U/L (at 37 °C)	IFCC primary reference method; Clin Chem Lab Med 2002;40:643-48	Haga Hospital, The Hague, The Netherlands
Consensus	Albumin	Alb	31–71 g/L	Consensus value = Mean laboratories	Not applicable
	Alkaline Phosphatase	AP	55–272 U/L (at 37 °C)		
	Phosphate	P	0.8–2.5 mmol/L		
	Urea	Urea	4.6–28.9 mmol/L		

The Dutch EQAS uses human, fresh frozen and commutable sera since 2005 [12–15]. Analytes in the EQA scheme are categorized into analytes for which reference measurement procedures were used to set target values (with subdivision for minerals, substrates, and enzymes, respectively; N = 13) and analytes for which no reference measurement procedures are available and for which consensus values are used as target values (consensus; N = 4). Categories as well as symbols listed here are used throughout the paper, especially in the figures. Reference Methods and Reference Laboratories involved with value assignment are listed.

Reports from about 220 participating laboratories are available a) bi-weekly after the deadline of each of the individual samples, b) quarterly aggregated per six samples and c) once a year as a condensed annual report reviewing all 24 samples in the annual cycle. Data in this paper are derived from the annual reports of the same clinical chemistry laboratories in 2005 and 2010. This led to the analysis of about 89760 overall data points and 5280 parameter-specific data points per year.

2.4. Supplemental files

Experimental proof of commutability of the human, liquid frozen EQA-materials used since 2005 is presented in a Supplemental data file. Proof of commutability of the liquid frozen EQA-materials is published for the 2005 EQA batch; for the subsequent batches it is assumed. Criteria for desirable bias, desirable precision and total allowable error are presented in a Supplemental Table 1.

2.5. Harmonizers

Beyond its EQA scheme the SKML also manufactures enzyme "harmonizers" with IFCC-values assigned by the enzyme reference lab in The Hague, the Netherlands. The harmonizers have been developed in the context of Calibration 2000, and should be considered as candidate reference materials [12]. The harmonizers enable clinical chemistry labs to calibrate serum enzymes (AST, ALT, LDH and γ-GT) [12]. Anno 2010, 130 out of 220 clinical chemistry laboratories use the enzyme harmonizers and belong to the Calibration 2000 method group.

2.6. Statistics

2.6.1. Regular EQA-samples

National EQA-data from 2005 and 2010 derived from twenty-four EQA-samples per year were analyzed. The obtained results were processed to evaluate equivalence among labs and to obtain insight into bias and precision trends for each analyte.

The first step in the processing of the data is done for each laboratory on a per analyte and per year basis using a linear regression model. Per analyte and per year for each laboratory a regression line is calculated through the 24 laboratory results as a function of the consensus method group mean values for each sample. The residual variance of the obtained line is an estimate of the intralaboratory variation. See Fig. 1. The intralaboratory bias is obtained from the mean of the differences of the 24 laboratory results from the reference values (or consensus value if no reference value is available). In the second step the intralaboratory variations and intralaboratory biases are used to obtain the overall imprecision, bias and interlaboratory variations.

- The overall imprecision is calculated as the median of the intralaboratory SD's, and expressed as a CV (percentage of the mean concentration of the samples)
- The overall bias is the median of the individual biases.
- The interlaboratory CV is calculated as the SD of the intralaboratory biases, and expressed as a CV (percentage of the mean concentration of the samples). Among labs, the median interlaboratory CV is a measure for the degree of harmonization of each analyte.

The SKML scores participating laboratories using an approach based on biological variability criteria derived from inter- and intraindividual variations [19,20]. For each lab the individual bias and intralaboratory CV is tested against the respective requirements for desirable bias and imprecision, and the combination is a prediction of the fraction of results that will be within allowable total error. In this approach lab performance is expressed as an allowable total error (TE_a) score. An individual lab complies with the TE_a score when the score is at least 95%. In addition, the percentage of labs that passes this TE_a score is calculated.

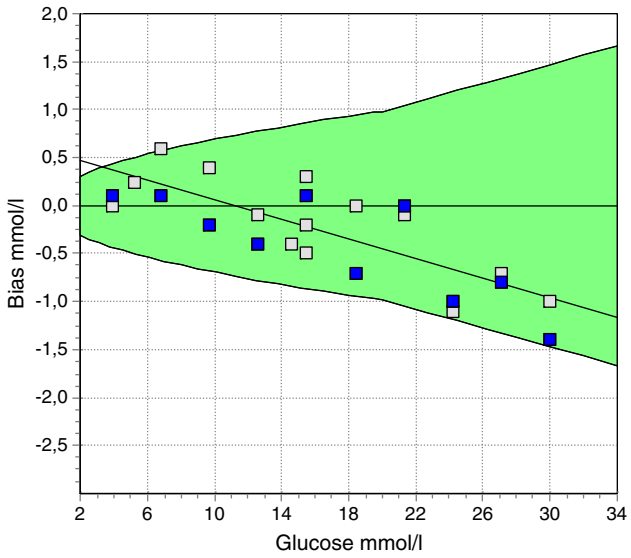

Fig. 1. Difference plot of glucose. The X-axis displays reference values; the Y-axis displays differences of measured results and reference values. Blue squares are the duplicates measured in the second half of the year (N = 12); grey squares are the duplicates measured in the first half of the year (N = 12). Intralab precision is calculated as the residual SD of the measurements around the regression line. The green area is the Total Allowable Error (TE_a) around the reference values.

All calculations have been performed at three aggregation levels: overall, per method group and per method. SKML has defined individual methods that mainly discriminate on analytical principle (e.g. in the case of serum creatinine the following analytical principles can be distinguished: endpoint Jaffe method; kinetic Jaffe method; compensated kinetic Jaffe method; enzymatic wet chemistry; enzymatic dry chemistry). Methods having the same analytical principle, are grouped in a method group (e.g. in the case of serum creatinine: a Jaffe method group and an enzymatic method group) and will share the same reference c.q. consensus value. Methods with significant differences in the chemical principle will have separate method groups with identical reference values but different consensus values. Graphs will use one of these three aggregation levels.

2.6.2. Spy-sample

To monitor commutability of the SKML EQA-pools with new methods and/or analyzers over the years, results of the native spy-sample and a regular EQA-sample with approximately the same analyte concentration are systematically compared (data not shown). Ideally the ratio should be 100% for all method groups and methods. To investigate whether the difference from 100% is statistically significant t-testing is performed.

3. Results

3.1. Equivalence of EQA-measurement results and evolution of bias and precision components

Fig. 2A shows the evolution of the interlaboratory CV from 2005 to 2010. In Fig. 2A the overall median interlaboratory CV in 2005 (x-axis) is plotted against the overall median interlaboratory CV in 2010 (y-axis). The figure can be interpreted in absolute terms, e.g. sodium (Na^+) has the lowest median interlaboratory CV in both years and the enzyme LDH has the highest median interlaboratory CV in 2005. The figure can also be interpreted in relative terms: the

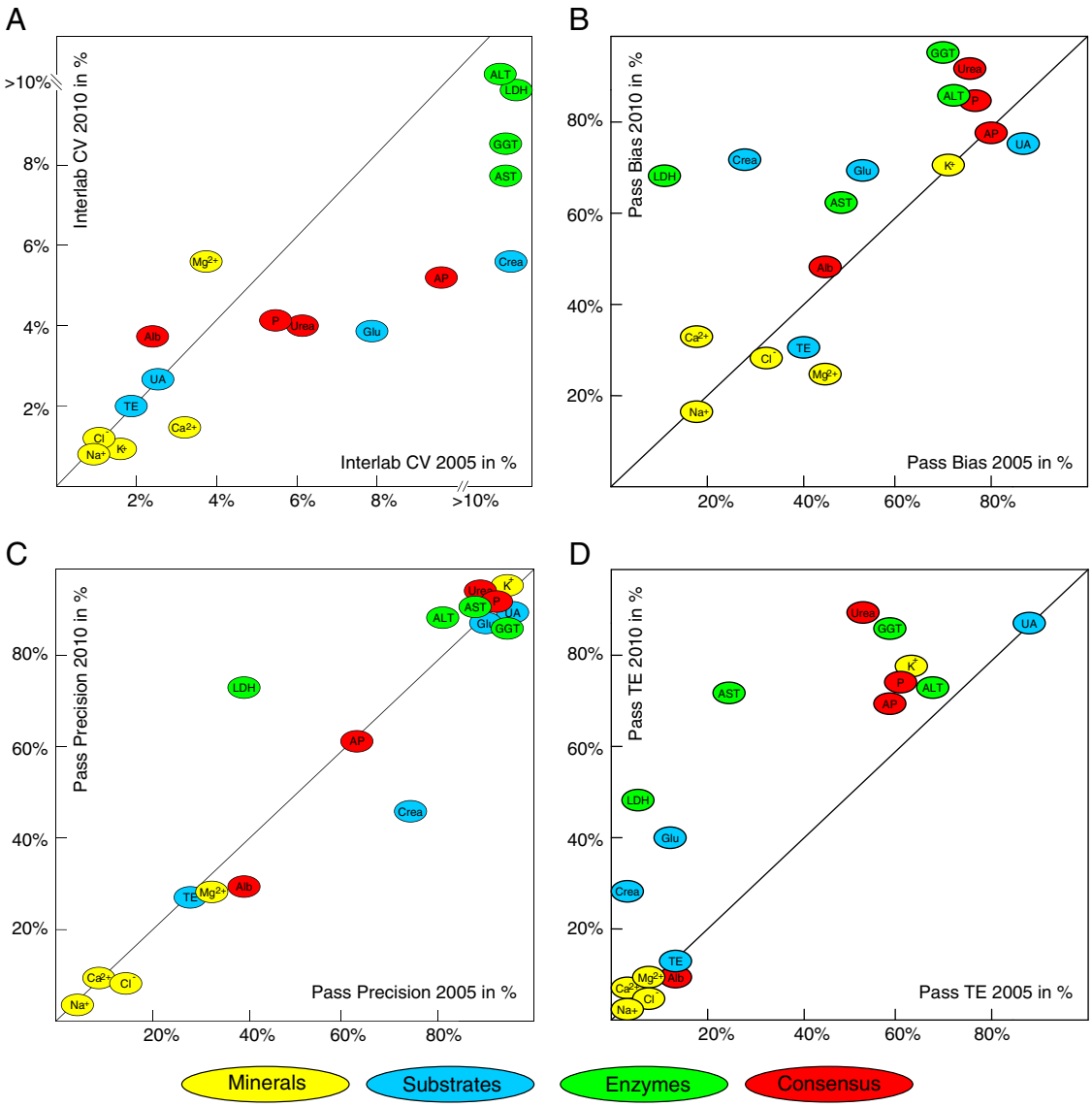

Fig. 2. Analytical performance trends for 17 general clinical chemistry parameters between 2005 and 2010. The analytes are divided in four color coded groups: three color coded groups refer to minerals, substrates and enzymes of which the EQA-materials have been value assigned with recognized reference methods (N = 13); the red colored group encompasses the parameters that rely on consensus means and have not been value assigned with recognized reference methods (N = 4).

- 2A: Evolution of the degree of Equivalence of test results
- 2B: Evolution of the Trueness Component between 2005 and 2010.
- 2C: Evolution of the Precision Component between 2005 and 2010
- 2D: Evolution of Total Allowable Error (TEa) between 2005 and 2010.

For legends of symbols: see Table 1.

median interlaboratory CV of analytes on the right side of the unity line, e.g. creatinine (Crea), has improved.

Interlaboratory CVs derive from contributions of bias (related to degree of standardization if applicable) and imprecision (related to intrinsic reproducibility of analytical methods). Change of the interlaboratory CV can thus derive from either changes in bias and/or precision. This is investigated in Fig. 2B and C. Fig. 2B shows the percentage labs passing the desirable bias in 2005 (x-axis) and 2010 (y-axis). Again interpretation can be in absolute terms: e.g. sodium (Na^+) has a low pass-rate and the enzyme γ-GT (GGT) has a high pass-rate, and in relative terms: the pass-rate for analytes on the left side of the line improved, e.g. for the enzyme LDH (LDH). Fig. 2C shows the percentage labs passing the desirable precision in 2005 (x-axis) and 2010 (y-axis). The precision Fig. 2C shows a neutral pattern: for some analytes the precision improved but for others the precision got worse. In Fig. 2C it is illustrated that overall pass rate for precision is more or less unchanged for most chemistry general analytes during the study period. Improved precision is noted for LDH, whereas deterioration occurred for creatinine. Interpretation of Fig. 2C in absolute terms shows consistently low pass-rates for sodium (Na^+) and high pass-rates for potassium (K^+); in relative terms the pass-rate of LDH improved but the pass-rate of creatinine (Crea) is in 2010 lower than in 2005 (45% against 76%). Fig. 2D presents the % of labs meeting the allowable total error in 2010 as compared to 2005. Fig. 2D illustrates that the % of labs meeting the allowable total error goal has improved in 2010 as compared to 2005 for 10 out of 17 parameters; the situation has not improved for seven parameters lying on the identity line. Especially the analytical performance of

the parameters Na^+, Cl^-, Mg^{2+}, Ca^{2+}, albumin and total protein is inadequate as compared to their corresponding allowable total error goals.

3.2. Monitoring standardization efforts

In Fig. 3A–B results of standardization efforts are given for four serum/plasma enzymes and for serum/plasma creatinine. Performance data are derived from the 2010 EQA-surveys. In Fig. 3A inter- and intralaboratory CV's in 2010 for serum/plasma enzymes are presented for the overall method group and for labs using the enzyme "harmonizer" [12]. Fig. 3B shows the pass rates for labs using enzymatic and Jaffe methods for creatinine, as well as for the overall method group.

3.3. Potential to improve equivalence with standardization/harmonization

In Fig. 4 harmonization potential is presented in 2010 for general clinical chemistry analytes (N = 17). Harmonization potential is expressed as the overall median interlaboratory CV / overall median intralaboratory CV ratio. The higher the ratio, the higher the remaining harmonization potential. The analytes are divided into four color groups: yellow, blue and green color coded groups refer respectively to minerals, substrates and enzymes of which the EQA-materials have been value assigned with recognized reference methods (N = 13); the red color coded group encompasses the parameters that rely on consensus means and have not been value assigned with recognized reference methods (N = 4).

In Fig. 4 lines are drawn for interlaboratory/intralaboratory CV ratios of 2.0 and 1.5 respectively. For analytes with a ratio > 2.0 there is, according to our experience, potential to improve equivalence with standardization/ harmonization efforts whereas standardization/ harmonization attempts cannot improve test equivalence for analytes with a ratio < 1.5.

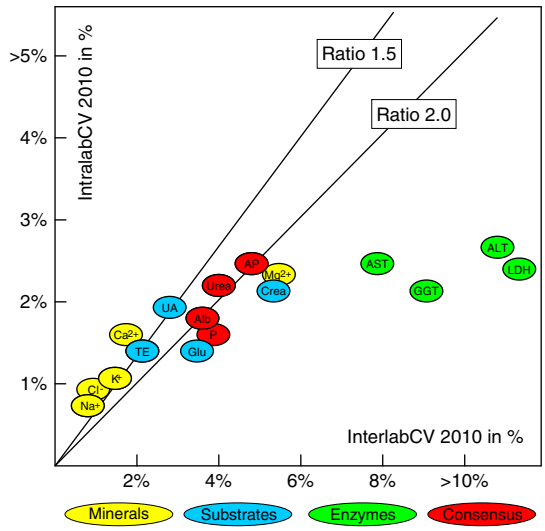

Fig. 4. Harmonization potential in 2010 for general clinical chemistry analytes (N = 17). For legends of symbols: see Table 1.

4. Discussion

Standardization/harmonization of medical laboratory tests is a prerequisite for producing globally interchangeable results. Interchangeability and equivalence of test results are essential for unequivocal interpretation of laboratory tests in clinical guidelines and clinical practice, and for worldwide comparability of lab results over space and time. To achieve this, many traceability milestones have been reached in the past two decades. Consequently, the clinical relevance of test equivalence through international standardization or harmonization of medical laboratory tests is broadly recognized by legal authorities (IVD 98/79/EC which became effective in Europe on 7 December 2003), metrological and reference institutes (www.bipm.org), professional organizations and commissions like the *Commission on Traceability in Laboratory Medicine* (C-TLM) of the *International Federation of Clinical Chemistry* (IFCC) (www.IFCC.org), the worldwide operating *Joint Committee on Traceability in Laboratory Medicine* (JCTLM) which is creating databases of certified reference laboratories, reference methods and reference materials (http://www.bipm.org/jctlm/), and the recent AACC Harmonization Initiative (www.harmonization.net).

The degree of test equivalence and the success of standardization / harmonization efforts can efficiently be monitored by EQA schemes that are based on commutable EQA-materials [4]. According to ISO/ REMCO N1129 commutability is a property of a reference or EQA-material, demonstrated by the equivalence of the mathematical

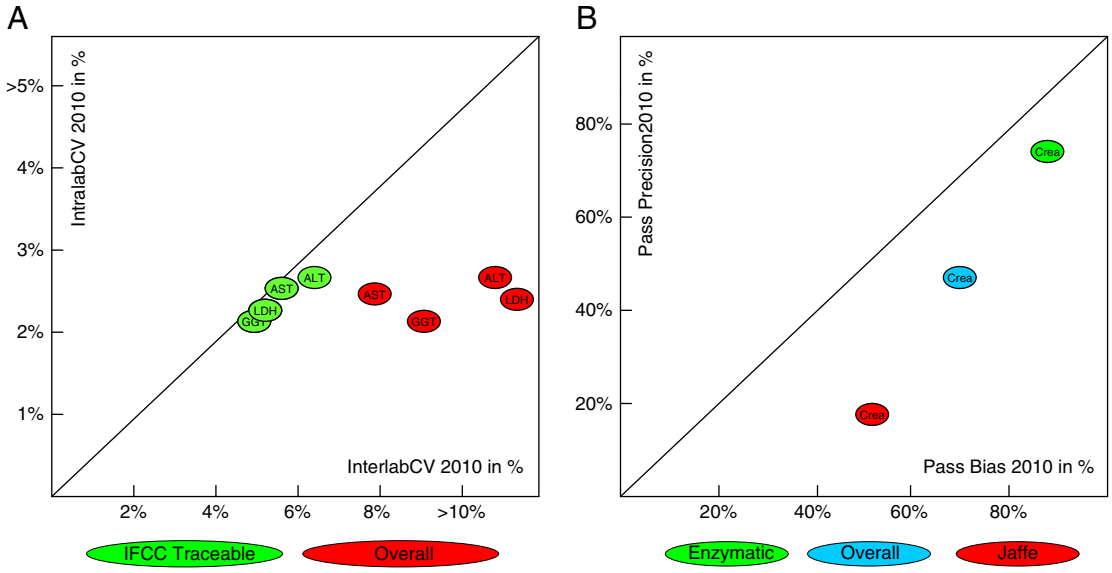

Fig. 3. Monitoring standardization efforts in 2010. Results of standardization efforts are given for four serum/plasma enzymes and for serum/plasma creatinine. 3A: Inter- and intralaboratory CV's in 2010 for serum/plasma enzymes; 3B: Pass rates for serum/plasma creatinine in 2010 for labs using enzymatic and Jaffe methods. For legends of symbols: see Table 1.

relationships among the results of different measurement procedures for the EQA-material and for representative samples of the type intended to be measured. In literature, we could not find EQA programs consistently using commutable specimens for monitoring quality of multiple general clinical chemistry parameters. The Dutch EQA, named the SKML (www.skml.nl) has such an EQA scheme in place (18, Supplemental data file). Since 2005 this EQA scheme is running for 17 general clinical chemistry parameters across all Dutch medical laboratories. According to the recent review of Miller et al. on EQA-evaluation capabilities, the Dutch EQA-scheme has an evaluation capability of the 1st category [4]. In this manuscript we summarize the degree of test equivalence in Dutch medical laboratories in 2010 as compared to the situation in 2005. Major findings from our EQA-surveys in the period 2005–2010 are the following.

4.1. Equivalence and evolution of bias and precision components

The best parameter to express equivalence of results (and thus the degree of standardization/harmonization) is the overall interlaboratory CV. In Fig. 2A, which shows the evolution of the interlaboratory CV from 2005 to 2010, it is illustrated that for 11 out of 17 parameters the interlaboratory CVs ameliorated in 2010 as compared to 2005. Yet, for albumin and Mg^{2+} the overall median interlaboratory CV increased, pointing to deterioration of quality and/or performance of the routine assays in 2010 as compared to 2005. In the bias Fig. 2B the majority of the analytes are on the left side of the line, which implies that overall bias has improved in 2010 as compared to the situation in 2005 for most general chemistry analytes. The greatest improvement has been reached for enzymes and creatinine; in case of Mg^{2+} the bias has deteriorated in 2010 as compared to 2005. From Fig. 2A-C it can be concluded that the improvement of the interlaboratory CV derives mainly from bias improvement by better standardization of the methods.

4.2. Monitoring standardization efforts

4.2.1. Enzymes

SKML with its Calibration 2000 program advocates that enzyme measurements should be standardized to IFCC Reference Measurement Procedures. To stimulate laboratories to do so, SKML supplies enzyme "harmonizers" with IFCC-values assigned by the enzyme reference lab in The Hague, the Netherlands [12]. The SKML EQA scheme monitors the success of this standardization effort and the aggregated results in Fig. 3A show that overall median interlaboratory CV's of labs following the advise (green ovals) are halved as compared to the overall method group (red ovals). The intralaboratory CV's of both groups are the same which means that the intrinsic quality of the methods (reproducibility) is the same and that the difference in equivalence is only due to standardization differences.

4.2.2. Creatinine

Opinion leaders advise the use of IDMS-standardized specific enzymatic methods for serum creatinine. This recommendation was adopted in the Dutch Calibration 2000 program. From Fig. 3B it can be derived that labs using unspecific Jaffe methods have a ~50% passing rate for bias and a ~20% passing rate for precision, whereas labs using more specific enzymatic methods have passing rates of ~90% and ~80%, respectively.

Both examples show that standardization/harmonization efforts can be monitored effectively with commutable, value-assigned EQA-materials and illustrate that this EQA-scheme displays the better methods and/or method groups.

4.2.3. Potential to improve equivalence with standardization

Equivalence of results from different laboratories is to an important degree determined by lack of bias (trueness) and precision. The contribution of bias is negligible when the interlaboratory CV of a method group of laboratories equals the intralaboratory CV. Yet, the more difference there is between inter- and intralaboratory CV, the higher the contribution of bias to the dispersion between lab results and the more effect of standardization/harmonization on improvement of equivalence can be expected. In Fig. 4 it is illustrated that overall there is much harmonization potential for the enzymes AP, ALT, AST, γ-GT and LDH; by using the Calibration 2000 "harmonizer", which enables IFCC-traceable standardization, the labs can significantly diminish their interlaboratory CV and bring the interlaboratory/intralaboratory ratio close to 1 (as illustrated in Fig. 3A).

4.2.4. Quality in absolute terms

The redesigned EQA program based on commutable specimens is an important tool for monitoring quality and accuracy of routine methods in clinical chemistry. From our study it can be seen that the pass-percentage of medical laboratories meeting the total allowable error has improved for 10 out of 17 parameters in 2010 as compared to 2005 (Fig. 2D), which can be explained by either bias reduction (Fig. 2B) or precision improvement (Fig. 2C). Secondly, for 15 out of 17 parameters the median interlaboratory CVs have improved or remained equal, but deterioration occurred for Mg^{2+} and albumin (Fig. 2A). In case of Mg^{2+} and albumin, quality and/or performance of the routine methods has declined triggered by workstation consolidation and shifting from e.g. laborious atomic absorption spectrometry respectively immunochemical methods to practical colorimetric methods. Thirdly, although reference systems for IFCC-standardization of enzymes are in place, interlaboratory CVs are still around 10%. In the EQA scheme we clearly demonstrated major improvement in interlaboratory CV in the Calibration 2000 method group using the enzyme "harmonizer" material, meaning that IVD-manufacturers should do a better job (Figs. 3A and 4). Finally, only for 6 out of 17 general clinical chemistry parameters desirable bias and precision criteria are met in ≈80% of the medical labs (figure not shown). In the case of serum creatinine the analytical performance is clearly method dependent: i.e., the national data from our EQA-surveys reveal that Jaffe methods have inferior analytical performance as compared to the enzymatic method group (Fig. 3B) [21]. This type of EQA-information can help lab professionals to choose the best method in their specific clinical setting.

5. Limitations

Commutability of the samples used in our EQA program was tested in the 2005 annual batch (Suppl data file). The same type of EQA samples have been used ever since. The assumption is that throughout the period 2005–2010 the EQA-samples prepared in subsequent batches were commutable as well. Yet, we realize that commutability can be a temporary property of EQA-samples in combination *with* certain measurement methods. Ideally, a new commutability study should be done per annual batch as throughout the years new routine methods and instruments come on the market and existing methods and instruments may be modified. The feasibility of annual commutability studies is questionable as this would be too much of a financial and logistic burden. To this end, the SKML has developed a pragmatic concept to keep an eye on commutability of annually prepared EQA-batches. Hitherto a spy-sample prepared according to NCCLS C37-A2 is included in the regular year EQA program for "sensing" commutability. Results of the native spy-sample and a regular EQA-sample with approximately the same analyte concentration are systematically compared. Commutability of EQA-samples with routine methods over the 2005–2010 period means that the EQA-sample behaves the same as the spy-sample across all routine methods. This implies that the ratio of EQA-sample and spy-sample should be the same among method groups. Ideally the ratio should be 100 % for all method groups and methods. But of course there is always a

random error. To investigate whether the difference from 100% is statistically significant t-tests are performed. If $t < 2.0$, there is no significant difference; if $t \geq 2$ there is a degree of non-commutability which would restrict the use of the EQA-material for trueness verification to certain parameters. Although not ideal nor being a replacement for a full commutability study, introduction of a spy-sample in the annual EQA-batches helps to give an indication of drifting commutability in subsequent annual EQA-batches in changing lab environments and analytical conditions. Future experiences should clarify the precise trigger role of the spy-material.

6. Conclusions

We conclude that the current Dutch EQA scheme for general clinical chemistry, based on the use of fresh frozen, commutable and mostly targeted EQA-materials, allows monitoring equivalence of test results, the effect of standardization efforts and analytical performance trends among laboratories over space and time. In addition, this EQA scheme helps to identify parameter deterioration. When targets are assigned with Reference Measurement Procedures, the commutable EQA-materials are in fact trueness verifiers, giving the lab professionals and EQA-organizers the tools to judge analytical bias. In addition, the scheme gives insight into the harmonization potential of the different analytes and/or method groups. We believe that EQAS organizers using commutable, targeted EQA-materials have a pivotal role in implementing, evaluating and ameliorating the metrological traceability concept. In combination with the scoring system, this toolbox gives lab professionals, IVD-industry, reference laboratories and EQAS-organizers insight into the current method performance as compared to the analytical performance needed for clinical use. Our data reveal opportunities for improvement for the different stakeholders. Even for general clinical chemistry analytes there is still a way to go.

Finally, the introduction of a native spy-material is a conceptual, keynote feature of our EQA-design for practical, affordable and yearly "sensing" of commutability of regular EQA-pools with current routine clinical chemistry methods.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cca.2012.09.027.

References

[1] Jansen RTP. The quest for comparability: Calibration 2000. Accredit Qual Assur 2000;5:363-6.
[2] Müller MM. Implementation of reference systems in laboratory medicine. Clin Chem 2000;46:1907-9.
[3] Miller WG, Myers GL, Gantzer ML, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57:1108-17.
[4] Miller WG, Jones GRD, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011;57:1670-80.
[5] Kallner A, McQueen M, Heuck C. Consensus agreement conference on strategies to set global quality specifications in laboratory medicine. Stockholm. April 24–26, 1999. Scand J Clin Lab Invest 1999;59:475-6.
[6] Ricós C, Ramón F, Salas A, Buño A, Calafell R, Morancho J. Minimum analytical quality specifications of inter-laboratory comparisons: agreement among Spanish EQAP organizers. Clin Chem Lab Med 2012;50:455-61.
[7] Gill J. Allowable limits of performance—statistical basis and implementation. Pathology 2011;43(S1):S17-20.
[8] Jansen RTP. Kalibratie 2000. Ned Tijdschr Klin Chem 1998;23:261-4.
[9] Baadenhuijsen H, Jansen RTP, Weykamp CW, Steigstra H, Kuypers AWHM. A nationwide split patient and control sample experiment to check the commutability of control samples [Abstract]. Clin Chem Lab Med 1999;37:S279 (abstract T250).
[10] Baadenhuijsen H, Scholten R, Willems HL, Weykamp CW, Jansen RTP. A model for harmonization of routine clinical chemistry results between clinical laboratories. Ann Clin Biochem 2000;37:330-7.
[11] Jansen RTP, Kuypers AWHM, Baadenhuijsen H, et al. Kalibratie 2000. Ned Tijdschr Klin Chem 2000;25:153-8.
[12] Jansen R, Schumann G, Baadenhuijsen H, et al. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum. An international study in the EC4 framework of the Calibration 2000 project. Clin Chim Acta 2006;368:160-7.
[13] van den Besselaar AMHP, Haas FJLM, Kuypers AWHM. Harmonisation of factor VIII:C assay results: study within the framework of the Dutch project 'Calibration 2000'. Br J Haematol 2006;132:75-9.
[14] van den Besselaar AMHP, Haas FJLM, van der Graaf F, Kuypers AWHM. Harmonization of fibrinogen assay results: study within the framework of the Dutch project 'Calibration 2000'. Int J Lab Hematol 2009;3:513-20.
[15] Jansen RTP, Baadenhuijsen H, Willems J. External Quality Assessment in the Netherlands: past, present and future. On the way to trueness assessment. In: van Pelt J, Poland DCW, editors. Clinical chemistry behind the Dykes. Ed. in honour of the 60th anniversary of the Netherlands Society for Clinical Chemistry and Laboratory Medicine. Verhagen Grafische Media; 2007. p. 87-97.
[16] Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, Jansen R. Commutability assessment of potential reference materials using a multicenter split-patient-sample between-field-methods (twin-study) design: study within the framework of the Dutch project "Calibration 2000". Clin Chem 2002;48:1520-5.
[17] Baadenhuijsen H, Weykamp C, Kuypers A, Franck P, Jansen R, Cobbaert C. Commuteerbaarheid van het huidige monstermateriaal in de SKML-rondzendingen van de algemene klinische chemie. Ned Tijdschr Klin Chem Labgeneesk 2008;33:154-7.
[18] Baadenhuijsen H, Kuypers A, Weykamp C, Cobbaert C, Jansen R. External quality assessment in the Netherlands: time to introduce commutable survey specimens. Lessons from the Dutch "Calibration 2000" project. Clin Chem Lab Med 2005;43:304-7.
[19] Fraser CG. General strategies to set quality specifications for reliability performance characteristics. Scand J Clin Lab Invest 1999;59:487-90.
[20] Ricos C, Alvarez V, Cava F, et al. Current databases on biologic variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491-500.
[21] Cobbaert C, Baadenhuijsen H, Weykamp CW. Prime time for enzymatic creatinine methods in pediatrics. Clin Chem 2009;55:549-58.