Clinica Chimica Acta 432 (2014) 82–89

Contents lists available at ScienceDirect

Clinica Chimica Acta

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

# External quality assurance programs as a tool for verifying standardization of measurement procedures: Pilot collaboration in Europe

C. Perich [a,b], C. Ricós [a,*], V. Alvarez [a,c], C. Biosca [a,d], B. Boned [a,e], F. Cava [a,f], M.V. Doménech [a,g], P. Fernández-Calle [a,h], P. Fernández-Fernández [a,i], J.V. García-Lario [a,j], J. Minchinela [a,k], M. Simón [a,l], R. Jansen [m]

[a] SEQC-Analytical Quality Commission, Barcelona, Spain
[b] Laboratori Clínic Bon Pastor, Barcelona, Spain
[c] Laboratori Clínic L'Hospitalet, Hospitalet de Llobregat, Spain
[d] Hospital Germans Trias i Pujol, Badalona, Spain
[e] Hospital Royo Villanova, Zaragoza, Spain
[f] Hospital Universitario Fundación Alcorcón, Madrid, Spain
[g] Laboratori Clínic Manso, Barcelona, Spain
[h] Hospital Universitario La Paz, Madrid, Spain
[i] Laboratoris Clínics Hospital Vall d'Hebron, Barcelona, Spain
[j] Hospital Universitario Virgen de las Nieves, Granada, Spain
[k] Laboratori Clínic Barcelonès Nord i Vallès Oriental, Badalona, Spain
[l] Consorci del Laboratori Intercomarcal de l'Alt Penedès, l'Anoia i el Garraf, Barcelona, Spain
[m] SKML Dutch Foundation for Quality Assessment in Medical Laboratories, Nijmegen, The Netherlands

ARTICLE INFO

*Article history:*
Received 13 August 2013
Received in revised form 22 October 2013
Accepted 8 November 2013
Available online 28 November 2013

*Keywords:*
Standardization
Harmonization
External quality assurance programs

ABSTRACT

*Introduction:* Current external quality assurance schemes have been classified into six categories, according to their ability to verify the degree of standardization of the participating measurement procedures. SKML (Netherlands) is a Category 1 EQA scheme (commutable EQA materials with values assigned by reference methods), whereas SEQC (Spain) is a Category 5 scheme (replicate analyses of non-commutable materials with no values assigned by reference methods).
*Aim:* The results obtained by a group of Spanish laboratories participating in a pilot study organized by SKML are examined, with the aim of pointing out the improvements over our current scheme that a Category 1 program could provide.
*Method:* Imprecision and bias are calculated for each analyte and laboratory, and compared with quality specifications derived from biological variation.
*Results:* Of the 26 analytes studied, 9 had results comparable with those from reference methods, and 10 analytes did not have comparable results. The remaining 7 analytes measured did not have available reference method values, and in these cases, comparison with the peer group showed comparable results. The reasons for disagreement in the second group can be summarized as: use of non-standard methods (IFCC without exogenous pyridoxal phosphate for AST and ALT, Jaffé kinetic at low-normal creatinine concentrations and with eGFR); non-commutability of the reference material used to assign values to the routine calibrator (calcium, magnesium and sodium); use of reference materials without established commutability instead of reference methods for AST and GGT, and lack of a systematic effort by manufacturers to harmonize results.
*Conclusions:* Results obtained in this work demonstrate the important role of external quality assurance programs using commutable materials with values assigned by reference methods to correctly monitor the standardization of laboratory tests with consequent minimization of risk to patients.


*Abbreviations:* AMP, 2-amino-2-methyl-1-propanol; CDC, Centers for Disease Control; DPD, 3,5-dichlorophenyldiazonium tetrafluoroborate; DTE, dithioerythritol; eGFR, estimated glomerular filtration rate; ERM, European Reference Material; GC–IDMS, gas chromatography–isotopic dilution with mass spectrometry; GLDH, glutamate dehydrogenase; ID-CP-MS, isotopic dilution with coupled plasma mass spectrometry; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine; IRMM, Institute for Reference Materials and Measurements (previously BCR, Bureau Comunitaire de Réference); NAC, N-acetylcysteine; NIST, National Institute for Standards and Technology; SRM, standard reference material; TPTZ, 2,4,6-tri-(2-pyridyl)-5-triazine.

* Corresponding author at: Plaza Gala Placídia 2, atico, 08006 Barcelona, Spain. Tel.: +34 93 218 14 20.
*E-mail address:* carmen.ricos@terra.com (C. Ricós).

0009-8981/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cca.2013.11.005

## 1. Introduction

It is well known that 70% of medical decisions are based on laboratory reports [1] and, consequently, it is of the upmost importance to achieve standardization among measurement procedures. This would also ensure that all laboratories produce harmonized results, with the aim of eliminating redundant requests and of facilitating interpretation of reports. With this in mind, laboratories should attain the following objectives: analytical accuracy, delivering reports in due time, efficacy and with a focus on patient safety [2].

Standardization can be defined as the property of obtaining comparable results, independent from analytical fundamentals, measurement method, and measurement procedure (the combination of instrument, reagent, calibrator and operative mode) [3]. Therefore, measurement results should be traceable to the International Units (SI) System and should satisfy the following conditions [4,5]:

- Routine procedures have to quantify exactly the same property as in the reference measurement and with the same specificity.
- A primary reference material or a recognized reference method should exist.
- Secondary calibrators, including routine calibrators should be traceable to the primary reference material or method, and should be commutable with patient samples.

In the case of non-compliance of one or more of these conditions, measurement procedures are not standardized but could be harmonized (or made comparable) if the results obtained were traceable to a reference or a consensus method when testing patient samples [2].

The tool that allows us to continuously verify compliance with these conditions is participation in interlaboratory comparison exercises. These exercises are formalized as External Quality Assurance (EQA) Programs and aim to evaluate not only analytical performance, but method performance, monitoring of in vitro diagnostics, continuous education, training and support [6].

Recently, EQA programs have been classified into six categories according to their ability to verify the degree of standardization or harmonization of the participating measurement procedures [7]:

1. Use of commutable EQA-samples, with values assigned by a reference method (RM), and with replicate analysis of the EQA-samples distributed.
2. Commutable EQA-samples with values assigned by a RM and without replicate analysis of EQA-samples.
3. Commutable EQA-samples, no RM values and with replicate analysis of EQA-samples.
4. Commutable EQA-samples, no RM values, no replicate analysis.
5. Non-commutable EQA-samples, no RM values and with replicate analysis.
6. Non-commutable EQA samples, no RM values and no replicate analysis.

EQA programs organized by the Spanish Society of Clinical Biochemistry and Molecular Pathology (SEQC) and running since 1976 use replicate analysis of the same non-commutable EQA-samples during each cycle (lasting one year), so they are in the 5th category. This implies that they can verify the concordance of each individual participant laboratory with other routine methods [8]. However, they cannot verify accuracy of the individual laboratory in a strict sense, because they do not distribute commutable materials (i.e. without matrix effects) and the materials are not traceable to reference standards. The Calibration Project 2000 in The Netherlands [9–11] and its external quality assurance scheme organizer SKML (*Stichting Kwaliteitsbewaking Medische Laboratorium Diagnostiek*) has achieved the most stringent conditions for several common chemistry analytes and is denoted as a Category 1 EQA scheme [7]. In particular, the materials for general chemistry are proven to be commutable and method target values are assessed by reference methods [12–14]. SKML provided SEQC with a set of its EQA materials, which were also distributed to usual SKML participants in the 2011 cycle as well as to some laboratories in the UK. This pilot study showed that the degree of standardization was better for Dutch laboratories than for the others [15].

The results obtained by the Spanish representatives are further examined here with the prospect of underlining the need for a category 1 EQA program to verify standardization in our setting. When non-standardization was found, the analytical procedures used were checked thoroughly to have an idea of the reasons for discrepancies.

## 2. Materials and methods

### 2.1. EQA materials

Six EQA samples from SKML were tested for 6 days (one sample per day in duplicate). These samples were human serum pools, frozen at −80 °C, distributed and maintained at this temperature until their analysis. Over the same time period (Jan to Jun 2013) SEQC-EQA samples were also analyzed.

### 2.2. Analytes tested

A total of 26 analytes were tested: alanine-aminotransferase (ALT), albumin, alkaline phosphatase, α-amylase (total), aspartate-aminotransferase (AST), bilirubin, calcium, chloride, cholesterol, creatine kinase (CK), creatinine, γ-glutamyltransferase (GGT), estimated glomerular filtration rate (eGFR), glucose, HDL-cholesterol, inorganic phosphate, iron, lactate dehydrogenase (LDH), lipase, magnesium, potassium, protein, sodium, triglyceride, urate and urea.

### 2.3. Laboratory platforms

Ten laboratories from different parts of Spain participated; two of them used two measurement procedures, giving a total of 12 procedures. Exceptionally, creatinine was measured by 13 procedures because one laboratory used the enzymatic method as a third procedure. The instruments used were: Abbott Architect ci 16200, Siemens Dimension Vista, Beckman Coulter Olympus AU 5400, Roche Cobas 711, Roche Cobas 6000 and Siemens Advia 2400.

### 2.4. Data analysis of imprecision and bias

#### 2.4.1. Within-batch imprecision

For each analyte and EQA-sample from SKML, imprecision was calculated from the duplicated analysis, applying the formula:

$$SD = \sqrt{(\sum d^2/2n)}$$

where d is the difference in duplicates (r1, r2) and n the number of pairs, 6 in this case. Generally $CV = SD / (\Sigma r / 2n)$.

#### 2.4.2. Between-laboratory imprecision

For each analyte and SKML-EQA-sample, the first value of each pair of results provided by each laboratory was taken to calculate the coefficient of variation among the 12 laboratory procedures and, then, the average among the six CV obtained for the six EQA-samples were calculated. The same procedure was applied to results of six SEQC-EQA samples analyzed during the pilot study period.

#### 2.4.3. Bias (systematic error)

For each analyte and EQA sample, the first result ($r_1$) of each laboratory procedure was taken to calculate the percentage deviation (PD) against the target value (TV), according to the formula:

$$PD = 100 * (r_1 - TV)/TV.$$

In this way, six PD values were obtained for each analyte and measurement procedure.

The target value was assigned by a reference method when available; if this kind of method did not exist, the target value was the peer group mean (laboratories using the same measurement method) obtained by the 200 participants in the SKML pilot study (170 labs from The Netherlands, 10 from Spain, 10 from Portugal and 10 from the UK).

Results obtained were considered to be comparable when the percentage deviation fell within the acceptability limit derived from biological variation for total error [16,17]. These limits were desirable, minimum or optimum [18,19] according to the criterion applied in the SEQC-EQA reports for each analyte. Because the six values throughout the concentration range are considered together, the overall view for each analyte designates the systematic error.

## 3. Results and discussion

### 3.1. Imprecision

Within and between-laboratory imprecision are shown in Table 1. Within-run laboratory imprecision falls within the acceptability limit derived from biological variation in all cases, except for sodium. For most of the analytes studied, imprecision between laboratories is greater when using SEQC-samples compared with SKML-samples.

### 3.2. EQA value assignment

Reference methods and reference laboratories assigning values to EQA materials, as well as routine methods and material or method used for traceability of routine calibrators are shown in Tables 2a–2c.

**Table 1**
Within and between laboratory imprecision for pilot study EQA samples.

| Analyte | $CV_{w\text{-}l}$ (%) | $CV_{b\text{-}l}$ (%) | |
|---|---|---|---|
| | | SKML | SEQC |
| Albumin | 0.6 | 3.3 | 4.2 |
| Alkaline phosphatase | 0.8 | 6.4 | 18.1 |
| ALT | 0.8 | 8.7 | 11.2 |
| α-Amylase (total) | 0.3 | 11.6 | 12.4 |
| AST | 1.2 | 6.0 | 5.1 |
| Bilirubin | 0.9 | 9.6 | 19.1 |
| Calcium | 0.5 | 3.5 | 3.7 |
| Chloride | 0.4 | 1.4 | 2.4 |
| Cholesterol | 1.0 | 4.1 | 3.5 |
| CK | 0.9 | 4.9 | 5.2 |
| Creatinine | 1.3 | 7.0 | 8.0 |
| GGT | 0.8 | 12.0 | 10.0 |
| Glucose | 0.7 | 5.9 | 3.1 |
| eGFR | 0.8 | 5.2 | 8.4 |
| HDL-cholesterol | 1.1 | 5.2 | 19.5 |
| Inorganic phosphate | 0.6 | 2.9 | 3.0 |
| Iron | 0.7 | 2.3 | 3.4 |
| LDH | 0.7 | 6.1 | 10.0 |
| Lipase | 0.6 | 4.1 | nd |
| Magnesium | 0.9 | 4.5 | 4.3 |
| Potassium | 0.5 | 1.5 | 2.0 |
| Protein | 1.2 | 3.2 | 5.3 |
| Sodium | 0.9 | 2.1 | 1.8 |
| Triglycerides | 0.6 | 3.5 | 4.2 |
| Urate | 0.6 | 5.2 | 5.4 |
| Urea | 0.9 | 3.4 | 4.1 |

nd, not measured.
$CV_{w\text{-}l}$: average within laboratory CV from the pilot study.
$CV_{b\text{-}l}$ from SKML: between laboratory CV from the pilot study.
$CV_{b\text{-}l}$ from SEQC: between laboratory CV of same laboratories for 6 SEQC–EQA samples distributed during the pilot study period.

**Table 2a**
Analytes tested with commutable EQA materials and values assigned by reference method obtaining comparable results.

| Analyte | Reference method and laboratory | Routine method (number of labs) | Routine calibrator traceability |
|---|---|---|---|
| Bilirubin | IFCC<br>DGKL, Hanover | DPD (9)<br>Jendrassik-Grof (2)<br>Vanadate oxidation (2) | NIST-SRM 916*<br>NIST-909b**<br>GC/MS Doumas |
| Chloride | Coulometry<br>INSTAND | Indirect potentiometry (12) | NIST-SRM 919*<br>NIST-SRM 909b**<br>NIST-SRM 956***<br>"Masterlot" |
| Cholesterol | CDC, Abell Kendall<br>Erasmus Medical Center, Rotterdam | Cholesterol-oxidase, esterase, peroxidase (12) | NIST-SRM 917b*<br>NIST-909b**<br>CDC IDMS |
| Creatine kinase | IFCC (NAC activator, 37 °C)<br>Haga Hospital, La Haya | UV, NAC activator (11)<br>UV, DTE activator (1) | ERM/IFCC-AD 455$<br>IFCC Absorptivity |
| HDL-Cholesterol | CDC designated comparison method<br>Erasmus Medical Center | Direct (10) | NIST-909b**<br>CDC |
| Glucose | GC–IDMS<br>INSTAND, Düsseldorf | Hexokinase (12) | NIST-SRM 917b*<br>NIST-SRM 965a***<br>IDMS Absorptivity |
| Potassium | ID-CP-MS<br>Flame spectrometry<br>INSTAND, Düsseldorf | Indirect potentiometry (12) | NIST-SRM 909b**<br>NIST-SRM 956***<br>"Masterlot" |
| Protein | Biuret<br>INSTAND, Düsseldorf | Biuret (12) | NIST-SRM 927c# |
| Urate | HPLC<br>Erasmus Medical Center | Uricase-colorimetric (3)<br>Uricase-peroxidase (9) | NIST-SRM 913*<br>NIST-SRM 909b**<br>IDMS |

DPD: 3,5-dichlorophenyldiazonium tetrafluoroborate; NAC: N-acetylcysteine; DTE: dithioerythritol; ERM: European Reference Material.
* Aqueous solution, pure substance.
** Lyophilized human serum with unproven commutability.
*** Frozen human serum with unproven commutability.
$ Lyophilized human serum with unproven commutability.
# Non-commutable bovine serum.

Tables 2a–2c are organized as follows with the analytes involved presented in alphabetical order:

a) Analytes tested with commutable EQA materials and values assigned by reference method obtaining comparable results.
b) Analytes tested with commutable EQA materials and values assigned by reference method not obtaining comparable results.
c) Analytes tested with commutable EQA materials without values assigned by reference methods but by peer groups.

### 3.3. Bias assessment

#### 3.3.1. Analytes tested with commutable EQA materials and values assigned by reference method obtaining comparable results

The analytes in this group are: bilirubin, chloride, cholesterol, HDL-cholesterol, creatine-kinase, glucose, potassium, protein and urate.

**Table 2b**
Analytes tested with commutable EQA materials and values assigned by reference method not obtaining comparable results.

| Analyte | Reference method and laboratory | Routine method (number of labs) | Routine calibrator traceability |
|---|---|---|---|
| ALT | IFCC (TRIS buffer with pyridoxal phosphate, 37 °C) | IFCC, with pyridoxal phosphate (4) | IFCC |
| AST | Haga Hospital, The Hague | IFCC without pyridoxal phosphate (8) | Absorptivity |
| α-Amylase | IFCC (Maltoheptaoside with p-nitrophenol and ethylidene, 37 °C) | Maltoheptaoside-p-nitro phenol and ethylidene (5) | IRMM/IFCC 456[@@] |
| | Haga Hospital, The Hague | Maltotrioside with 2 chloro-p-nitrophenol (3) | "Masterlot" IFCC Absorptivity |
| Calcium | Atomic absorption spectrometry | Arsenazo (7) | NIST-SRM 915[*] |
| | INSTAND, Düsseldorf | O-cresolphthalein (5) | NIST-SRM 909b[**] NIST-SRM 956[***] |
| Creatinine | GC–IDMS | Enzymatic (1) | NIST-SRM 914[*] |
| eGFR | DGKL, Hanover | Jaffé kinetic (5) | NIST-SRM 909b[**] |
| | | Jaffé kinetic compensated (7) | NIST-SRM 967[****] IDMS[&] |
| GGT | IFCC (gamma-glutamyl-3-carboxy-4-nitroanilide > 4 mmol/L, 37 °C) Haga Hospital, The Hague | IFCC (12) | ERM/IFCC 452[&&] |
| LDH | IFCC (lactate to pyruvate, 37 °C) Haga Hospital, The Hague | Lactate to pyruvate (7) Pyruvate to lactate (5) | IRMM 453[@@] |
| Magnesium | Atomic absorption spectrometry | Arsenazo (1) | NIST-SRM 929[*] |
| | INSTAND, Düsseldorf | Xylidyl blue (3) Chlorophosphonazo (1) | NIST-SRM 909b[**] |
| Sodium | Flame emission spectrometry | Indirect potentiometry (12) | NIST-SRM 909b[**] |
| | INSTAND, Düsseldorf | | NIST-SRM 956[***] |

eGFR, estimated glomerular filtration rate.
[*] Aqueous solution, pure substance.
[**] Lyophilized human serum with unproven commutability.
[***] Frozen human serum with unproven commutability.
[****] Frozen human serum with proven commutability.
[&] Animal tissue, commutable.
[&&] Animal tissue l, not commutable.
[@@] Human tissue, commutability not proved.

For each analyte, the percentage deviation of results obtained versus the corresponding reference method value is compared with the quality specification for total error derived from biological variation, being the desirable limit for 5 analytes (cholesterol, HDL-cholesterol, glucose, potassium and urate), the minimum limit for two analytes (chloride and protein) and the optimum limit for two analytes (bilirubin and creatine kinase). More than 95% of deviations obtained are within the limits, so we can consider that results provided by participant laboratories are comparable and, consequently, method procedures used are well standardized. This happens both when the same measurement procedure is used in all laboratories (6 analytes) and when different measurement procedure are used (3 analytes), even with different traceability of routine calibrators (Table 2a).

In the case of glucose, only two results are not comparable, corresponding to the two measurement procedures using a calibrator physically traceable to the NIST-SRM 917 reference material, which has an aqueous matrix and, consequently, is expected to be non-commutable with human serum.

**Table 2c**
Analytes testing commutable EQA materials without values assigned by reference methods but by peer groups.

| Analyte | Comparison method | Routine method (number of labs) | Routine calibrator traceability |
|---|---|---|---|
| Albumin | Peer group mean | Bromocresol green (10) Bromocresol purple (2) | ERM-DA470k[**] NIST-SRM 927c[#] |
| Alkaline phosphatase | Overall mean | 4-nitrophenyl-phosphate, AMP buffer (12) | IFCC Internal calibrator Absorptivity |
| Inorganic phosphate | Overall mean | Phosphomolybdate, 340 nm (12) | NIST-SRM 21861 Internal calibrator |
| Iron | Overall mean | Ferrozine (3) TPTZ (4) Ferene (1) Colorimetric (2) | NIST-SRM 937[*] |
| Lipase | Overall mean | 1,2 diglyceride with glycerol-kinase and glycerol-3-phosphate-peroxidase (5) | Internal calibrator |
| Triglyceride | Overall mean | Lipase/glycerol kinase (11) | IDMS[***] NIST-SRM 909b[**] Absorptivity |
| Urea | Overall mean | Urease-GLDH (12) | IDMS NIST-SRM 912[*] NIST-SRM 909b[**] |

AMP: 2-amino-2-methyl-1-propanol; TPTZ: 2,4,6-tri-(2-pyridyl)-5-triazine; GLDH: Glutamate dehydrogenase.
[*] Aqueous solution, pure substance.
[**] Lyophilized human serum with unproven commutability.
[***] Frozen human serum with unproven commutability.
[#] Bovine serum, not commutable.

### *3.3.2. Analytes tested with commutable EQA materials and values assigned by reference method not obtaining comparable results*

This group includes: alanine aminotransferase (ALT), aspartate aminotransferase (AST), α-amylase (total), calcium, creatinine, eGFR, γ-glutamyltransferase (GGT), lactate dehydrogenase (LDH), magnesium and sodium.

The percentage deviation of results obtained versus the corresponding reference method value is depicted in Figs. 1–3 for analytes with clear reasons for non-comparability. Percentage deviations are compared with quality specifications for total error derived from biological variation, with the minimum limit for two analytes (calcium and sodium), the optimum limit for two analytes (ALT and GGT) and the desirable limit for the remaining six analytes.

*3.3.2.1. ALT and AST.* 4 out of the 12 participants use a routine method based on the IFCC reference method (TRIS buffer with pyridoxal-5-phosphate (P5P), at 37 °C), and 8 labs use a similar method but without P5P. Percentage deviation of results versus the reference method value

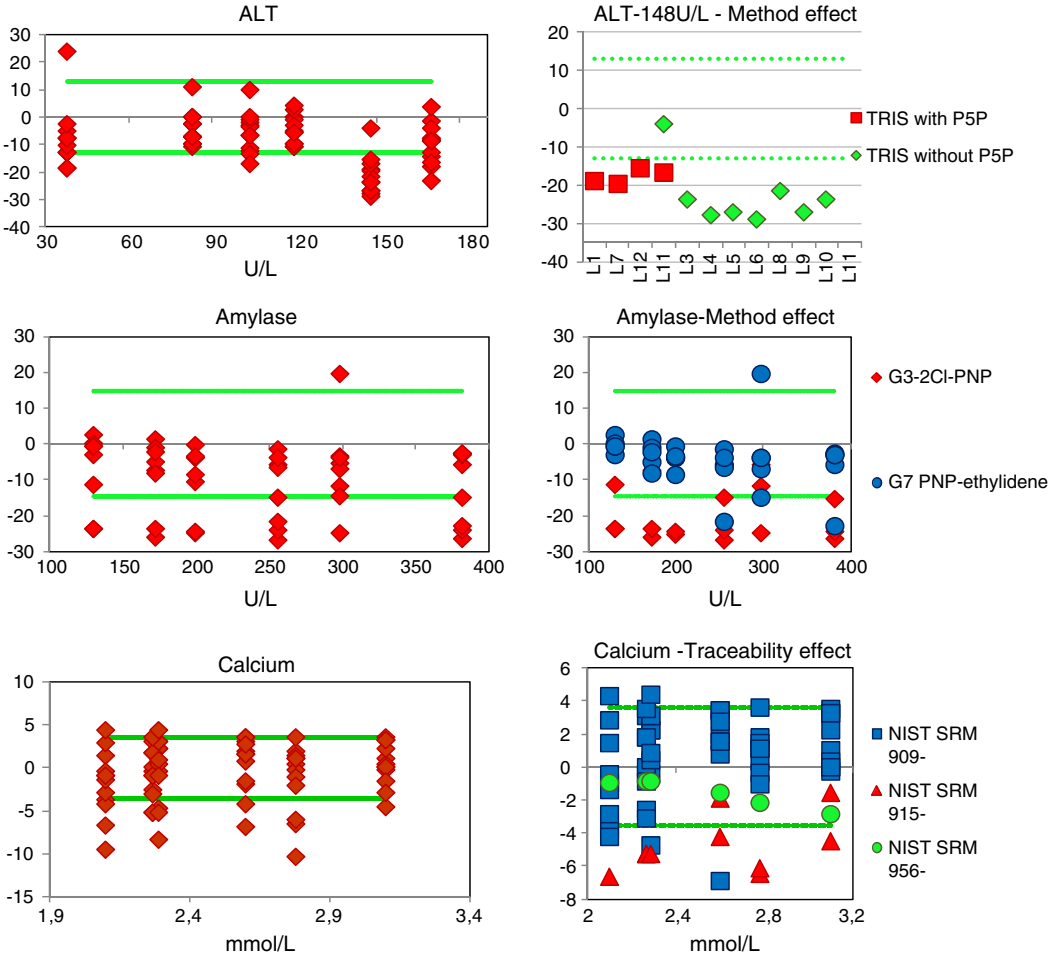


**Fig. 1.** Analytes tested with commutable EQA materials and values assigned by reference method not obtaining comparable results. Y-axis: Percentage deviation compared with the reference method value. X-axis in all figures except ALT at 148 U/L: reference method value for the six EQA-samples (ALT: 38, 83, 104, 120, 148, 170 U/L; amylase: 130, 172, 199, 256, 298, 382 U/L; calcium: 2.10, 2.27, 2.29, 2.60, 2.78, 3.10 mmol/L); Laboratory procedures 1 to 12; dashed lines represent acceptability limits derived from biological variation (optimum for ALT, desirable for amylase, minimum for calcium).

fall outside the optimum limits for both analytes in a number of cases, as was shown before by Jansen et al. [20]. For example ALT at 148 U/L shows lower results for measurement procedures without P5P (Fig. 1), which may be due to a possible instability of ALT in frozen samples, as described by Infusino et al. [21]. However, studies on the SKML samples showed stability at −80 °C for ALT and AST with recoveries against the reference target value falling around 100% for all instrument groups, both in samples measured in January and in samples tested in December, year after year. When considering the traceability of the routine calibrator used, no discrepancies are seen among the 11 measurement procedures traceable to the IFCC reference method and the single procedure using molar absorptivity.

*3.3.2.2. α-Amylase (total).* Only 8 out of the 12 participants test this analyte, 5 of them using the G7-PNP ethylidene method (IFCC recommended) and 3 the G3-2chloride-PNP. There are several results below the acceptability limit for the whole concentration interval studied (Fig. 1). When stratifying the results according to the measurement method, we can see that substrate heptaoside method achieves satisfactory results, whereas substrate trioside obtains lower results, also with a high interlaboratory imprecision as shown in Table 1. The results are in agreement with those previously reported [20]. No effect due to the calibrator traceability is observed for this analyte.

*3.3.2.3. Calcium.* The majority of results obtained fall within the minimum limits derived from biological variation for total error, independently of the measurement procedure used. However, there are some unacceptably low values provided by laboratories using a routine calibrator traceable to the NIST-SRM 915 standard (calcium carbonate aqueous dissolution, a matrix so different to human serum that it is considered to be non-commutable) (Fig. 1). Consequently, it seems that there is a direct relationship between the accuracy of results and the matrix of the standard used to assign the values of the routine calibrators.

*3.3.2.4. Creatinine.* Results fall, in general, within the acceptable interval. However, at the level of clinical relevance (79 μmol/L) several high results can be seen. Fig. 2 shows the deviation of participant measurement procedures classified according to recommendations of the SEQC-Renal Function Commission [22]. Correct results are provided by the enzymatic procedure and, also by 5 of the 7 compensated Jaffe kinetic measurement procedures; the other 5 procedures based on the non-compensated Jaffé kinetic method give unacceptably high results. Similar findings were published by Delanghe et al. [23]. When stratifying results according to the metrological traceability of routine calibrators used, we can see that all procedures traceable to IDMS give correct results but not those traceable to HPLC. Moreover, when the routine calibrator is traceable to a reference material, no clear association with the accuracy of results was observed.

*3.3.2.5. Estimated glomerular filtration rate (eGFR).* eGFR is calculated from the creatinine result assuming that samples come from a 55-year-old white female. The formulae used are MDRD or MDRD–IDMS

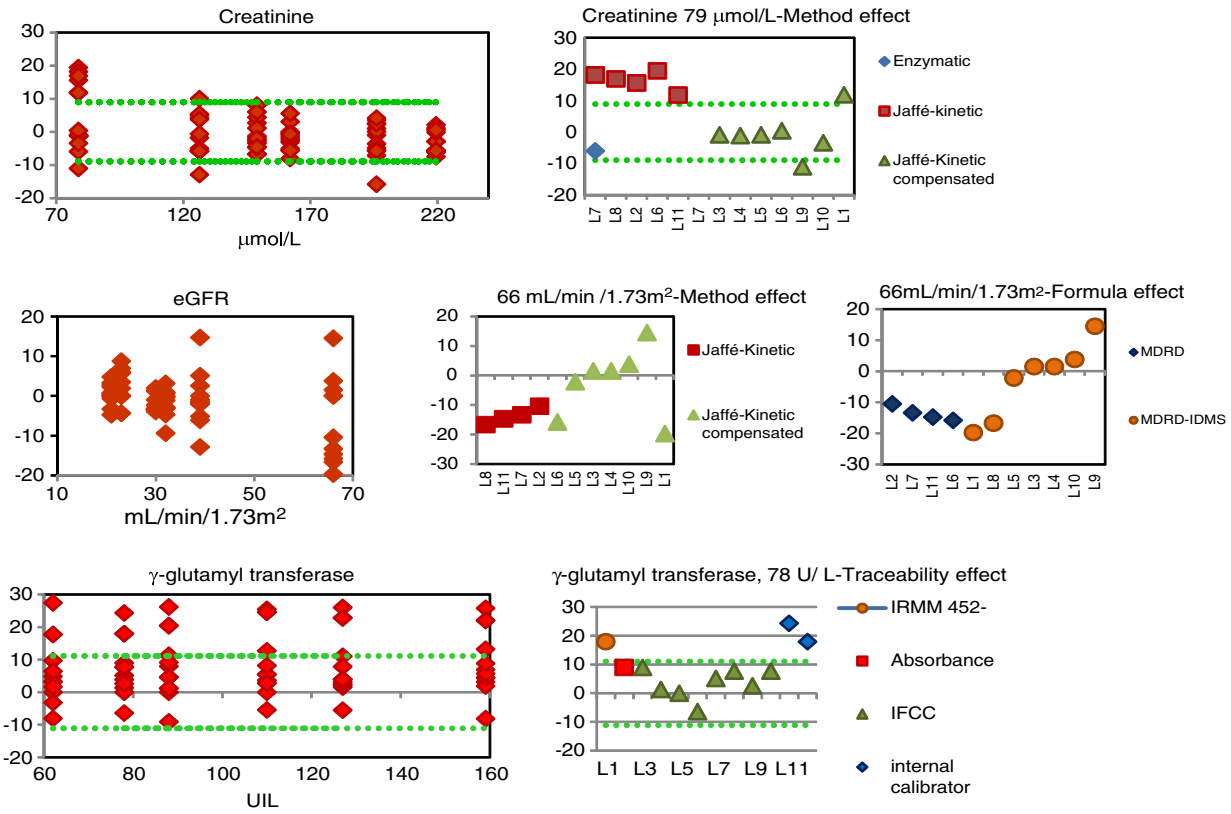


**Fig. 2.** Analytes tested with commutable EQA materials and values assigned by reference method not obtaining comparable results. Y-axis: Percentage deviation compared with the reference method value. X-axis in all figures except creatinine at 79 µmol/L, eGFR at 66 mL/min/1.72 $m^2$ and GGT at 78 U/L: reference method value for the six EQA-samples (creatinine:79, 126, 149, 162, 196, 219 µmol/L; eGFR: 21, 23, 30, 32, 39, 66 mL/min/1.72 $m^2$; GGT: 62, 78, 88, 110, 127, 159 U/L); Laboratory-procedures 1 to 12; dashed lines represent acceptability limits derived from biological variation (desirable for creatinine, optimum for GGT).

and the reference material or methods used to offer traceability are described in the previous paragraph. Results obtained are depicted in Fig. 2, where we can see deviations up to 15%, in both senses at rates of 40 and 66 mL/min/1.73 $m^2$. When stratifying results obtained at the level of maximum clinical interest (66 mL/min/1.73 $m^2$) according to the creatinine method used, it is evident that all results from the Jaffé kinetic method have a negative deviation, with falsely low results under the cut-off value of 60 mL/min/1.73 $m^2$. Such results could lead clinicians to incorrectly assess the severity of risk to health in patients. When stratifying results according to the formula employed we can see that neither of the two equations compensate for the low values located in the area of risk to health (negative deviation).

Concerning traceability, we can see that procedures traceable to the NIST 967 overstate patient risk when using the Jaffé kinetic method because results have a negative bias, whereas results are more satisfactory when using the compensated Jaffé procedure.

*3.3.2.6. GGT.* Results obtained are quite different from each other and some of them exceed the acceptable limit, as was observed before [20]; also, a high interlaboratory imprecision can be seen in Table 1. This happens despite the fact that all laboratories use the same metrological method.

Stratifying results according to the metrological traceability of the calibrator values proves that the higher values correspond to one procedure traceable to the IRMM 452 standard material and two laboratories using an internal calibrator (masterpool from the manufacturer). On the other hand, calibrator values traceable to a reference method give results well within the acceptability limits.

*3.3.2.7. LDH.* The participant laboratories use two different methods:

- Lactate to pyruvate (L–P), using the same substrate as for the reference method (7 laboratories).
- Pyruvate to lactate (P–L), (5 laboratories).

Because results from the second method are approximately double those of the first, bias of the two groups is studied separately. The substrate pyruvate procedure, as with the IFCC reference method, obtains deviations within the acceptable limits, except for two results corresponding to one laboratory using a molar absorptivity. The lack of standardization seen here is in agreement with previous reports [15,20].

*3.3.2.8. Magnesium.* Participating laboratories use three different methods (Table 2b). Percentage deviation of results versus the reference method falls generally within the acceptable limits; however, various exceptions are distinguished that cannot be associated with the metrological procedure used. All procedures are traceable to non-commutable reference materials (Table 2b) and maybe this fact could explain the poor results observed.

*3.3.2.9. Sodium.* Only a limited percentage of results fall within the acceptable limits. There are several values under the lower limit, coming from the metrological procedure traceable to the NIST-SRM 909b, which is not a proven commutable material.

*3.3.3. Analytes testing commutable EQA materials without values assigned by reference methods but by peer groups*

This section includes: albumin, alkaline phosphatase, inorganic phosphate, iron, lipase, triglyceride and urea.

Percentage deviation of results versus the peer group mean is compared with quality specifications for total error derived from biological variation. The limits are minimum for albumin, desirable for alkaline phosphatase, inorganic phosphate, urea, and optimum for iron, lipase and triglyceride.

*3.3.3.1. Albumin.* Two different metrological procedures, bromocresol green (BCG) and bromocresol purple (BCP), are used to test albumin with 2.5% low biased results from BCP compared with BCG, comparing each laboratory with its specific peer group. Practically all procedures using BCG are traceable to the ERM-DA 470 reference material, and all using BCP are traceable to the NIST-SRM 927c. The percentage deviations compared with the corresponding peer group mean fall within the acceptable limit in the majority of cases. One laboratory using BCG and a routine calibrator traceable to ERM-DA 470 obtained low results over the whole concentration range studied.

*3.3.3.2. Other analytes.* When testing alkaline phosphatase, inorganic phosphate, lipase, triglyceride and urea, all participant metrological procedures are based on the same method and, therefore, are compared with the overall mean for each analyte.

Although traceability of routine calibrators is different, percentage deviation of results versus the overall mean fall within the acceptable limits, a good harmonization among users of the same method when testing these analytes.

Lastly, in the case of iron procedures used in this study, all have the same metrological basis (reduction of $Fe^{3+}$ to $Fe^{2+}$ and colorimetric reading of the complex formed) and, for this reason, the target value is the overall mean. Percentage deviations obtained reach the optimum acceptability limit, so a good harmonization for iron procedures is demonstrated.

## 4. General discussion

The results show that one of the most likely explanations for unacceptable performance is the lack of commutability of the reference material used to assign the value of a routine calibrator. Detailed information about the procedure used to assign routine calibrator values is often not provided by manufacturers, and laboratories cannot know if the traceability chain has been entirely maintained. For this reason, several authors recommend using reference methods instead of reference materials to assign the value of the routine calibrators, because commutability of reference materials with human serum for the different routine procedures cannot always be ensured [2,24]. Also it was shown in this work that the lack of commutability of the routine calibrators may be a reason for non-standardization [2,25–29], because the physical traceability chain is broken. An example is with calcium testing, with the possibility of obtaining false low results when the measurement procedure is traceable to a non-tested commutable reference material. This constitutes a serious problem, as without this information and despite operating according to the manufacturer's specifications, the laboratory could produce false results. As manufacturers are obliged by the European In Vitro Diagnostics Directive [30] to provide the customer with this information, any error of this kind is the manufacturer's responsibility.

Failure of standardization may have an important repercussion on patient healthcare. A clear example is the non-compensated Jaffé kinetic procedures and those not traceable to IDMS, which produce high results at a concentration close to the clinical decision level and a false number of eGFR results below the cut-off point. In this way, erroneous diagnosis of chronic renal disease could happen, with the consequent harm for patients and increase of healthcare costs. The use of the different formulae for estimation of GFR, based on routine calibrator traceability might not be enough to harmonize results at the clinical decision level [31].

This lack of harmonization has been recently studied by Panteghini, who considers the need to urgently revise clinical decision limits in a shared task with laboratory professionals, IVD manufacturers, EQA organizers and clinicians [32]. Whereas in this pilot study use of commutable EQAs has never given different results among instruments using the same method, the same cannot be said for the SEQC-EQA with non-commutable EQAs where important discrepancies among instruments using the same method can be seen [33]. This is another reason that obliges us to move towards a higher category of EQA.

In summary, results obtained in this work demonstrate the important role of external quality assurance programs using commutable EQAs with values assigned by reference methods to correctly monitor performance in laboratory medicine and, consequently, to correctly identify patient risk. The main inconvenience for its implementation is the economic cost. This difficulty could be minimized by integrating this type of EQA material with limited frequency in one of the cycles of the current program. Our organization is firmly committed to promoting this activity within our country.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cca.2013.11.005.

## References

[1] Report of the review of NHS pathology services in England. An independent review for the Department of Health. London: Department of Health; 2006:6.
[2] Miller GW, Myers GL, Gantzer ML, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2012;57:1108–17.
[3] Siekmann L. Metrological traceability — a concept for standardization in laboratory medicine. Clin Chem Lab Med 2013;51:953–7.
[4] Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. Clin Biochem 2009;42:236–40.
[5] Asociación Española de Normalización y Certificación. In vitro diagnosis medical devices — measurement of quantities in biological samples — metrological traceability of values assigned to calibrators and control materials UNE-EN ISO 17511; 2004.
[6] Maziotta D, Harel D, Schumann G, Libeer JC. Guidelines for the requirements for the competence of EQAP organizers in medical laboratories. International Federation of Clinical Chemistry (IFCC)/Education and Management Division (EMD)/Committee of Analytical Quality (C-AQ); 2002.
[7] Miller GW, Jones GRD, Horowitz GL, Weycamp C. Proficiency testing/external quality assessment current challenges and future directions. Clin Chem 2011;57:1670–80.
[8] Ramón F, Alsina MJ, Álvarez V, et al. XXXII Programa de Garantía Externa de la Calidad de Bioquímica (suero) de la Sociedad Española de Bioquímica Clínica y Patología Molecular. In: http://www.contcal.org; 2011. [Accessed 3-11-12].
[9] Jansen RTP. The quest for comparability: Calibration 2000. Accred Qual Assur 2000;5:363–6.
[10] Jansen RTP, Kuypers AWHM, Baadenhuijsen H, et al. Kalibratie 2000. Ned Tijdschr Klin Chem 2000;25:153–8.
[11] 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.
[12] 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.
[13] 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.
[14] Cobbaert C, Weykamp C, Franck P, 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.
[15] Jansen R, Jassam N, Thomas A, 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 2013. http://dx.doi.org/10.1016/j.cca.2013.11.003.
[16] Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Consensus agreement. In: Hyltoft Petersen P, Fraser CG, Kallner A, Kenny D, editors. Strategies to set global analytical quality specifications in laboratory medicine. Scan J Clin Lab Invest, 59(7); 1999. p. 585.
[17] Minchinela J, Ricós C, Perich C, et al. Biological variation database, and quality specifications for imprecision, bias and total error (desirable, optimum and minimum). The 2012 update. http://www.westgard.com/biodatabase1/htm. [accessed 26-11-2012].
[18] Fraser CG, Hyltoft Petersen P, Libeer JC, Ricós C. Proposals for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8–12.
[19] SEQC, Comité de Garantía de la Calidad, Comisión de Calidad Analítica. Recomendación para el uso de las especificaciones de la calidad analítica notificadas

en el informe mensual de un programa de garantía externa de la calidad de la SEQC. http://www.seqc.es/es/Comisiones/7/51/102.
[20] 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.
[21] Infusino I, Schumann G, Ceriotti F, Panteghini M. Standardization in clinical enzymology: a challenge for the theory of metrological traceability. Clin Chem Lab Med 2010;48:301–7.
[22] Comisión de Función Renal de la SEQC. Ecuaciones de estimación del filtrado glomerular a utilizar según combinación procedimiento de medida-analizador. http://www.seqc.es/es/Comisiones/18/22/279/Noticias_y_Anunciosde_la_Comision_de_Funcion_renal_|_Comité_Cientifico/ (accessed 26-11-2012).
[23] Delanghe JR, Cobbaert C, Harmoinen A, Jansen R, Laitinen P, Panteghini M. Focusing on the clinical impact of standardization of creatinine measurements: a report by the EFCC Working Group on Creatinine Standardization. Clin Chem Lab Med 2011;49:977–82.
[24] Cattozzo G, Guerra E, Ceriotti F, Franzini C. Commutable calibrator with value assigned by the IFCC reference procedure to harmonize serum lactate dehydrogenase activity results measured by 2 different methods. Clin Chem 2008;54:1349–55.
[25] Canalías F, García E, Sánchez M. Metrological traceability of values for α-amylase catalytic concentration assigned to a commutable calibrator materials. Clin Chim Acta 2010;411:7–12.
[26] Miller WG, Myers GL, Rej R. Why commutability matters. ClinChem 2006;52:553–4.
[27] Vesper HW, Miller WG, Myers GR. Reference materials and commutability. Clin Biochem Rev 2007;28:139–47.
[28] CLSI. Characterization and qualification of commutable reference materials for laboratory medicine. Approved guideline. CLSI document C53-A. Wayne, (PA). CLSI; 2010.
[29] Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998;31:449–57.
[30] Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Official Journal of the European Communities of 7 December 1998:L331.
[31] Alvarez V, Ricós C, Fernández-Calle P, et al. Impacto de la trazabilidad de los métodos de creatinina en la estimación del filtrado glomerular. Abstract al VI Congreso Nacional de Laboratorio Clínico. Barcelona; Octubre 2012 https://intranet.pacifico-eetings.com/amsysweb/publicacionOnline.jsf?id=95 (accessed 26-11-2012).
[32] Panteghini M. Implementation of standardization in clinical practice: not always an easy task. Clin Chem Lab Med 2012;50:1237–41.
[33] SEQC, Comité de Garantía de la Calidad. XXXIII programa de Garantía externa de la Calidad de Bioquímica (suero) de la Sociedad Española de Bioquímica Clínica y Patología Molecular. http://www.contcal.org/qcweb/Documents/40%20Avaluacio%20anual/50%20Programes%202012/40.%20Suero.pdf; 2012 (accessed 01-06-2013).