EurA1c Trial Group

# $EurA_{1c}$: a review of 9 years performance of up to 5120 laboratories in 29 country and 26 manufacturer groups according to current and proposed criteria of the IFCC Model for Quality Targets in relation to the use of $HbA_{1c}$ for monitoring and diagnosis of diabetes

https://doi.org/10.1515/cclm-2026-0276
Received February 19, 2026; accepted March 31, 2026;
published online May 19, 2026

**Abstract**

**Objectives:** A major objective of the IFCC Committee for Education in the Use of Biomarkers in Diabetes is to monitor the success of global standardization of $HbA_{1c}$ through benchmark evaluation of the performance by country and manufacturer.

**Methods:** From 2016 to 2024 annual trial rounds were organized. Fresh whole blood and lyophilized hemolysate specimens were sent to 29 External Quality Assessment organizers and distributed to up to 5,120 laboratories. Mean results from 2016 to 2024 and from 2024 were evaluated by country and manufacturer, according to current (5 mmol/mol−0.46 %) and proposed (3 mmol/mol−0.28 %) criteria of the IFCC Model for Quality Targets.

**Results:** The state of the art in 2024 is an overall bias of 0.6 mmol/mol (0.05 % in NGSP units) with a between laboratory coefficient of variation of 4.0 % (2.7 % in NGSP units). This analytical performance implies that overall 2 % of the laboratories failed to meet the current and 16 % the proposed criterion. Differentiation to country and manufacturer groups showed that 0–10 % of the countries failed to meet the current and 8–35 % the proposed criterion. Of the manufacturers 0–30 % failed to meet the current and 1–58 % the proposed criterion. Results of previous years 2016–2024 were similar. Lyophilized hemolysate specimens were not commutable for some methods.

**Conclusions:** The performance of $HbA_{1c}$ measurement has been stable over the last decade, but substantial differences between country and manufacturer groups were observed. Many met the current criterion, but the picture was mixed when proposing tighter criteria. Whenever possible organizers should use fresh whole blood specimens.

**Keywords:** $HbA_{1c}$; collaborative EQA; EQA organizers; current and proposed IFCC criteria for monitoring and diagnosis of diabetes

## Introduction

Hemoglobin $A_{1c}$ is an essential analyte for effective diabetes monitoring and diagnosis [1]. To ensure quality of $HbA_{1c}$ tests, the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) developed a global standardization system based on the concept of metrological traceability, over two decades ago [2, 3].

To demonstrate the success of the implementation and efficacy of the standardization in daily practice, the IFCC Committee for Education in the Use of Biomarkers in Diabetes (C-EUBD) organizes the annual $EurA_{1c}$ trial. This is a benchmark study in collaboration with 29 External Quality Assessment (EQA) organizers representing up to 5,000 laboratories.

The ultimate goal of standardization is to achieve a performance level that meets requirements derived from the clinical needs of laboratory tests. In 2015 the C-EUBD set a criterion of a total allowable error (TAE) for $HbA_{1c}$ of 5 mmol/mol (0.46 % in units of the National Glycohemoglobin Standardization Program, NGSP) at 50 mmol/mol (6.7 %) [4]. This criterion, composed of bias and between laboratory coefficient of variation (BLCV), forms the basis for the current evaluation of the $EurA_{1c}$ trial. It has been one decade since the quality targets for $HbA_{1c}$ were set by the IFCC and it is now timely to assess if they remain fit for purpose or if they need to be revised to meet current clinical needs derived

**Corresponding author: Cas Weykamp**, MCA Laboratory, European Reference Laboratory, Queen Beatrix Hospital, Beatrixpark 1, Winterswijk, 7101 BN, The Netherlands, E-mail: c.w.weykamp@skbwinterswijk.nl

from the use of Hb$A_{1c}$. The present criterion of 5 mmol/mol (0.46 %), was based on the assumption that treatment might be changed when Hb$A_{1c}$ changes at least 5 mmol/mol (0.46 %) at two consecutive points in time. However, this is not stringent enough for use of Hb$A_{1c}$ for the diagnosis of diabetes. Therefore, the C-EUBD considers to update the criterion. An accepted approach when setting quality targets for diagnosis is to base them on the biological variation of the analyte. A reliable source of robust quality criteria is the biological variation database of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). Using these data the TAE criterion for minimum performance is 2.5 mmol/mol (0.23 %) at 50 mmol/mol (6.7 %) [5]. As decimal places are not reported with SI units, this is rounded up to 3 mmol/mol (0.3 %). We used the Eur$A_{1c}$ data to investigate the effect of tightening the criterion on country and manufacturer performance and to determine how many laboratories would meet the present criterion of 5 mmol/mol (0.5 %) and the proposed revised criterion of 3 mmol/mol (0.3 %).

The first round of the trial was organized in 2016 [6]. Since then, the number of participating laboratories has doubled, and data of nine rounds have become available. The aim of the present paper is to describe the trend in performance over the years, including the present state of the art. Bias, BLCV, and the failure rates to meet the current and proposed performance criteria of the C-EUBD are evaluated to see if Hb$A_{1c}$ is fit to be used for monitoring and the diagnosis of diabetes. Overall results are presented, but are also shown differentiated in groups to identify poorly performing countries and manufacturers. The aim of publishing the results is to stimulate the poor performers to improve their quality to achieve the best patient care.

# Materials and methods

## Study design

The study design has remained unchanged since its inception in 2016 [6]. An overview of the 2024 round is shown in Figure 1. From a pool of fresh whole blood donations (violet), specimens of fresh whole blood (WB; green) and lyophilized hemolysate (LH; amber) were prepared. EQA organizers generally prefer to use WB specimens as these specimens better reflect samples used in routine clinical practice, but when logistic constraints prevented this, LH specimens were provided. Specimens were shipped in bulk to up to 29 EQA providers in 25 countries, who in turn sent these samples to up to 5,120 laboratories for analysis. Target values were assigned with the IFCC Reference Measurement Procedure (IFCC RMP; yellow). Homogeneity and stability (grey) were tested according to ISO 13528 standards and criteria were met. The data of nine annual rounds were evaluated in three ways: overall, differentiated by country, and differentiated by manufacturer. For each group, the degree to which current and proposed tighter criteria were met, as well as the analytical parameters bias and BLCV were considered. Non-commutability, a potential disadvantage of LH specimens, was also investigated. The trial was approved by the Ethic Committee and donors signed an informed consent form.

## Sample preparation, logistics and assigned values

In each trial, 250 mL donations of whole blood were collected into Potassium Ethylenediaminetetraacetic acid ($K_2$EDTA) donation bags from diabetic and non-diabetic volunteers and used to make two pools with Hb$A_{1c}$ concentrations at a pre-diabetes level of approximately 42 mmol/mol (6.0 %) and at a diabetic level of approximately 58 mmol/mol (7.5 %). From each pool WB and LH specimens were made. Logistics were organized in such a way that participating laboratories assayed the WB specimens within five days of donation. LH specimens were manufactured immediately after donation and shipped and analyzed within 1–6 months.

Target values were assigned with the IFCC-RMP [2] by five approved IFCC network laboratories; each laboratory measured the specimens in 4-fold. The 9-year mean of the specimens with the pre-diabetes Hb$A_{1c}$ concentration was 41.9 mmol/mol (5.98 %) with a mean expanded uncertainty of 0.7 mmol/mol (0.06 %); range 0.6–0.8 mmol/mol (0.05–0.07 %). The 9-year mean of the specimens with the diabetic Hb$A_{1c}$ concentration was 57.7 mmol/mol (7.43 %) with a mean expanded uncertainty of 0.9 mmol/mol (0.08 %); range 0.8–1.1 mmol/mol (0.07–0.10 %).

## Definitions

– Following the consensus statement of the American Diabetes Association, European Association for the Study of Diabetes, IFCC, and the International Diabetes Federation on the global standardization of Hb$A_{1c}$ measurement, Hb$A_{1c}$ concentrations are expressed in SI units (mmol/mol) with NGSP units (%) in parentheses [7]; BLCVs are also reported as two units (NGSP in parentheses) because they are different according to the unit system used [8].

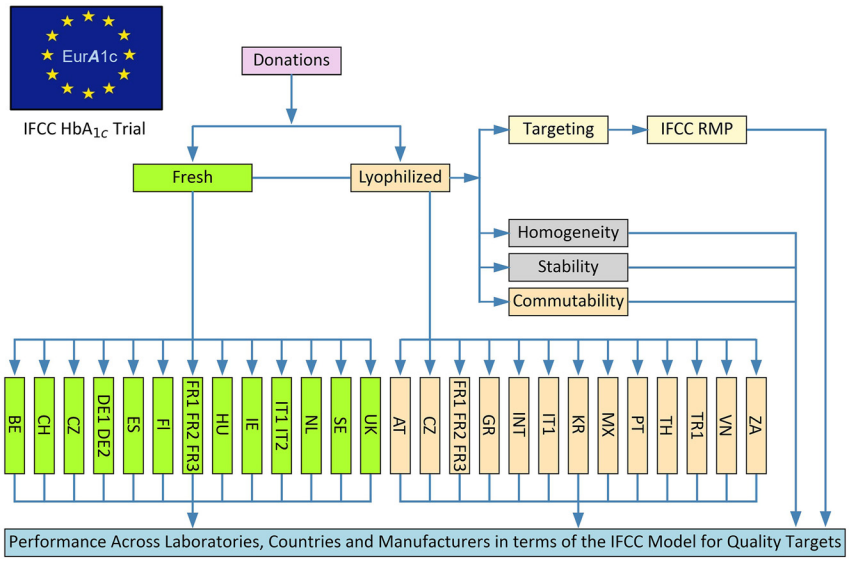


**Figure 1:** Design of the IFCC EurA1c Trial 2024. Donations (violet) from which WB (green) and LH (amber) specimens are prepared and used in the respective countries (blue); supporting tests (grey). Countries: Austria ÖQUASTA (AT), Belgium Sciensano (BE), Czech Republic SEKK s.r.o (CZ), Finland labquality (FI), France CTCB (FR1), France ProBioQual (FR2), France Asqualab (FR3), Germany INSTAND (DE1), Germany SPMD-RfB (DE2), Greece ESEAP (GR), Hungary QualiCont (HU), ERL International group (INT), Ireland IEQAS (IE), Italy CRB (IT1), Italy CRRVEQ (IT2), Korea, KEQAS (KR). Mexico Grupo ERL (MX), Netherlands SKML (NL), Portugal PNAEQ-INSA (PT), South Africa NHLS (ZA), Spain SEQC$^{ML}$ (now SEMEDLAB) (ES), Sweden Equalis (SE), Switzerland CSCQ (CH), Thailand National Institute of Health (TH), Turkey TUBITAK UME (TR), United Kingdom Weqas (UK), Vietnam QCC (VN).

- In this paper the word “methods” is used for end-user test systems as supplied by manufacturers (other synonyms are measurement procedures and platforms).
- The evaluation under the heading “manufacturers” deals with the evaluation of results of groups of laboratories using the same method. It is recognized that a manufacturer may have more than one method or method type; therefore, for each manufacturer their respective methods are listed as separate manufacturers, e.g. Tosoh G8 and Tosoh G11 are listed as separate manufacturers.
- The evaluation under the heading “countries” deals with the evaluation of results of groups of laboratories of the respective EQA organizers. It is recognized that some EQA organizers may have participants from countries outside the country of the EQA organizer. This might be relevant, as national regulations regarding analytical performance for HbA$_{1c}$ can differ and potentially influence laboratory performance. Until now data on this are not collected but this may be given attention in future EurA$_{1c}$ rounds.
- The current criterion of the IFCC C-EUBD, set in 2015, is 5 mmol/mol (0.46 %). A group of laboratories met this criterion when at least 95 % of the laboratories of that group had a TAE not exceeding this criterion [4].
- The proposed criterion, considered by the IFCC C-EUBD 10 years after the current criterion was set, is 3 mmol/mol (0.28 %). A group of laboratories met this criterion when at least 95 % of the laboratories of that group had a TAE not exceeding this criteria.

## Data collection

EQA organizers collated the results from their participants and forwarded them to the IFCC network coordinator. Outliers, arbitrarily defined as a value outside +/− 25 % of the target, were removed. Outliers made up 0.7 % of results in WB specimens and to 1.4 % in LH specimens.

## Presentation

Due to the large scale of the trial, over nine years, choices on what, how and where to present results were necessary. In this paper, data are presented as condensed results in tables and figures in the main body and with more detail in the Supplementary Materials. The evaluation was done with the mean of the high and low specimens for the 2024-trial round (state of the art) and the mean of the nine 2016–2024 rounds (trend analysis). Thus 2016–2024 refers to EurA$_{1c}$ rounds undertaken in the years 2016–2024, 2024 relates specifically to the round undertaken in 2024. The complete data, including detailed results per manufacturer per country, are available on the trial website [9].

## Selection for evaluation

Results of EQA organizers only included those that joined the EurA$_{1c}$ trial at least twice in the past five years and had at least 15 participating laboratories in each of these years.

Results of manufacturers only included well defined (e.g. Roche cobas c 513) platforms for which there were at least 15 participating laboratories per year in the last two years.

Non-commutability of LH specimens was evaluated for methods in which at least 80 results were available for both WB and LH specimens.

## Statistical analysis

Results were evaluated according to the IFCC Model for Quality Targets (IFCC-MQT) [4]. The model is based on the concept of total error which includes the principal sources of analytical error: bias and imprecision. Bias is plotted on the vertical axis with scaling in SI units (mmol/mol) and NGSP units (%). Imprecision, expressed as BLCV (also in SI – and NGSP units) is on the horizontal axis. The current IFCC criterion was set at 5 mmol/mol (0.46 %) at the 2σ level and applies to HbA$_{1c}$ concentrations of 50 mmol/mol (6.7 %). In the graph, this criterion is shown as a line drawn from plus five and minus 5 mmol/mol (+/−0.46 %) on the vertical axis to 5 % (3.4 %) on the horizontal axis. A result within the triangle (represented by a dot) meets the current criterion. The model can be applied at the level of a single laboratory (imprecision is defined as the within-laboratory CV) or for groups of laboratories (imprecision is the BLCV). The latter is used in this paper to evaluate the performance of specific countries and manufacturer groups.

Bias is defined in equation (1), where M1 and M2 are the mean measured HbA$_{1c}$ concentrations in the specimens 1 and 2 of each round of the trial, and T1 and T2 the target values of the respective specimens 1 and 2 assigned with the IFCC RMP.

$$\text{Bias} = \frac{(M1 - T1) + (M2 - T2)}{2} \tag{1}$$

BLCV is defined as the mean of the BLCV in samples 1 and 2 of each round of the trial.

Expanded uncertainties are derived solely from the standard deviation of the measurement results with a coverage factor 2.

The percentage of laboratories failing the current criterion of 5 mmol/mol (0.46 %) and the proposed criterion of 3 mmol/mol (0.28 %) was derived from bias and BLCV according to our previous paper [10]. First, probabilities in standard deviation units were calculated. The probability of exceeding a criterion at the high concentration end is defined in Eq. (2), and the probability of exceeding a criterion at the low concentration end is defined in Eq. (3).

$$P_{\text{high}} = \frac{(C - B)}{0.01*(V*50)} \tag{2}$$

$$P_{\text{low}} = \frac{(C + B)}{0.01*(V*50)} \tag{3}$$

$C$ is the criterion (thus five or 3 mmol/mol), $B$ is the bias and $V$ is the BLCV. Second, the probability in standard deviation is converted to the probabilities in percentages P %$_{\text{high}}$ and P%$_{\text{low}}$. Third, the total probability of exceeding a criterion is the sum of P%$_{\text{high}}$ and P%$_{\text{low}}$.

Example

In 2024, the overall bias (B) was + 0.6 mmol/mol and the overall BLCV (V) was 4.0 % (Table 2).

Then at the proposed criterion of 3 mmol/mol:

$$P_{\text{high}} = \frac{(3 - 0.6)}{0.01*(4*50)} = 1.2\,\text{SD}$$

$$P_{\text{low}} = \frac{(3 + 0.6)}{0.01*(4*50)} = 1.8\,\text{SD}$$

Standard Deviations correspond to 12 % one-sided probability of exceedance=P%$_{\text{high}}$.

Standard Deviations correspond to 4 % one-sided probability of exceedance=P%$_{\text{high}}$.

The total probability of exceeding the criterion=12 % + 4 %=16 %.

Non-commutability of LH specimens (also described as the matrix effect) was calculated using equation (4) and (5).

$$M_B = (B_I - B_W) \tag{4}$$

Here M$_B$ is the matrix effect on the bias, B$_w$ is the bias for the WB specimens and B$_l$ the bias for the LH specimens.

$$M_V = (V_L - V_W) \tag{5}$$

Here M$_V$ is the matrix effect on the BLCV, V$_W$ is the BLCV for the WB specimens and V$_L$ the BLCV for the LH specimens.

# Results

## Overall

Table 1 shows an overview of the number of participating EQA organizers and laboratories and the overall bias and BLCV throughout the years for the users of both WB and LH specimens.

Table 2 shows the overall performance in analytical terms and failure rates of laboratories according to the current and proposed IFCC criteria. For example: in 2024, the mean bias of all laboratories that used WB specimens was +0.6 mmol/mol (+0.05 %) with a BLCV of 4.0 % (2.7 %).

**Table 1:** Number of participants and overall bias and between-laboratory CV throughout the years 2016–2024.

| General data | | | Fresh whole blood samples | | | | | | Lyophilized hemolysate samples | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | N-Pt[b] | % Labs using WB | N-EQ[c] | N-Pt[b] | Bias | | CV | | N-EQ[c] | N-Pt[b] | Bias | | CV | |
| | | | | | IFCC m/m[d] | NGSP % | IFCC % | NGSP % | | | IFCC m/m[d] | NGSP % | IFCC % | NGSP % |
| 2016 | 2,166 | 70 | 10 | 1,517 | 0.2 | 0.02 | 4.4 | 3.0 | 10 | 649 | −0.5 | −0.05 | 4.9 | 3.3 |
| 2017 | 2,647 | 68 | 12 | 1,809 | 0.7[a] | 0.06 | 4.1 | 2.8 | 11 | 838 | 0.5 | 0.05 | 5.3 | 3.6 |
| 2018 | 3,983 | 72 | 15 | 2,878 | 0.3 | 0.03 | 4.6 | 3.1 | 13 | 1,105 | 0.4 | 0.04 | 5.8 | 3.9 |
| 2019 | 4,575 | 66 | 16 | 3,038 | 0.4 | 0.04 | 4.6 | 3.1 | 13 | 1,537 | 1.1[a] | 0.10 | 6.2 | 4.2 |
| 2020 | 5,120 | 64 | 17 | 3,286 | 0.6 | 0.05 | 4.7 | 3.2 | 15 | 1,834 | 1.0[a] | 0.09 | 6.9 | 4.7 |
| 2021 | 4,077 | 62 | 14 | 2,524 | 0.9[a] | 0.08 | 4.1 | 2.8 | 14 | 1,553 | 0.5 | 0.05 | 5.9 | 4.0 |
| 2022 | 4,325 | 64 | 14 | 2,788 | 0.4 | 0.04 | 4.1 | 2.8 | 14 | 1,537 | 0.0 | 0.00 | 6.1 | 4.1 |
| 2023 | 4,082 | 62 | 13 | 2,546 | −0.1 | −0.01 | 3.9 | 2.7 | 15 | 1,536 | −0.2 | −0.02 | 6.1 | 4.1 |
| 2024 | 4,284 | 64 | 16 | 2,753 | 0.6 | 0.05 | 4.0 | 2.7 | 15 | 1,531 | 0.1 | 0.01 | 5.5 | 3.7 |
| Mean | 3,918 | 66 | 14 | 2,571 | 0.4 | 0.04 | 4.3 | 2.9 | 13 | 1,346 | 0.3 | 0.03 | 5.9 | 4.0 |

[a]Bias higher than the expanded uncertainty of the assigned value. [b]Number partipants. [c]Number EQA organizers. [d]mmol/mol.

This level of analytical performance meant that 2 % of the laboratories did not meet the current, and 16 % did not meet the proposed IFCC criterion.

## Results by EQA organizer (country)

Table 3 presents the performance per country (EQA Organizer) in analytical terms (SI units) and failure rates of laboratories according to the current and proposed IFCC criteria. Figure 2 visualizes per country data when the current and proposed criteria of the IFCC-MQT are applied. The left-hand side presents the average of the results from 2016 to 2024 with the panels on the right hand side focused on data from the 2024 round. In panel A, the country codes are plotted in the ovals. In the panels below (B, C) the country codes are replaced by the percentages of laboratories in those countries that did not meet the current (B) or proposed (C) criterion. For example, in panel A, laboratories in Belgium (BE) had a bias near zero and a BLCV of just over 3 % (2 % NGSP). In the panels below, this translates into the percentages of Belgian laboratories that did not meet the current criterion (<1 % in panel B) and proposed criterion (8 % in panel C). Similarly, on the right-hand side of Figure 2, the results of the most recent round, are shown in the panels D, E, and F. Details of number of participants, bias, and BLCV for WB and LH specimens in the respective years 2016–2024 are in the Supplementary Tables 1–3.

## Results by manufacturer

Similar to the results by country, Table 4 shows the relation between analytical performance and percentages of laboratories failing the current and proposed criteria, for manufacturers. Figure 3 mirrors the structure of Figure 2. In the upper panels, the dots indicate the manufacturers' respective methods with a code, and in the panels below, the percentages of laboratories of those manufacturers that did not meet the current and proposed criteria. For example, in panel A, laboratories using the Roche cobas c 513 method (Q), had a bias of +0.9 mmol/mol (+0.08 %) and a BLCV of 2.4 % (1.6 %). In the panels below this translates into percentages of cobas c 513 users that did not meet the current criterion (<1 % in panel B) and proposed criterion (4 % in panel C). Similarly, on the right hand side of Figure 3, the results of the most recent round, are shown in the panels D, E, and F. Details of number of participants, bias, and BLCV for WB and LH specimens in the respective years 2016–2024 are in the Supplementary Tables 4–6; in these tables all manufacturers with at least five users were included.

Data on the market share of the manufacturers and the analytical principles are in Supplementary Table 7.

## Commutability of lyophilized specimens

Table 5 presents the data of the non-commutability of LH specimens expressed as the mean matrix effect on bias and BLCV in the trials from 2016 to 2024. Methods are grouped

**Table 2:** Overall performance in analytical terms and failure rates of laboratories according to the current and proposed IFCC criteria.

| Perspective | Parameter | Mean 2016–2024 | | 2024 | |
|---|---|---|---|---|---|
| | | IFCC | NGSP | IFCC | NGSP |
| Analytical performance | Bias | +0.4 | +0.04 % | +0.6 | +0.05 % |
| | Between laboratory CV | 4.3 % | 2.9 % | 4.0 % | 2.7 % |
| % Laboratories Failing IFCC criteria | Current criterion of 5 mmol/mol | 3 % | 3 % | 2 % | 2 % |
| | Proposed criterion of 3 mmol/mol | 18 % | 18 % | 16 % | 16 % |

per analytical principle. Example: for the Bio-Rad D-10 method, there were 607 results available for WB specimens and 614 for LH specimens. The mean bias for WB was 1.2 mmol/mol (0.11 %) and 0.5 mmol/mol (0.05 %) for LH. Then the matrix effect is 0.5–1.2= −0.7 mmol/mol (0.06 %). Its expanded uncertainty is 0.3 mmol/mol (0.03 %), which implies that the matrix effect is statistically significant. Details on the matrix effect per year are in the Supplementary Table 8. Calculations on the correction for non-commutability of LH specimens are in Supplementary Table 9.

# Discussion

## Overall

Table 2 shows that overall, 98 % of the laboratories met the current IFCC criterion in 2024 vs. a mean of 97 % in the years 2016–2024. The proposed tighter criterion was met by 84 % of the laboratories in 2024 and 82 % in 2016–2024. This indicates that performance has remained stable over the last nine years. The performance goal of the IFCC C-EUBD, derived from the IFCC MQT, is that 95 % of the laboratories meet the criterion [4]. The IFCC goal for groups is achieved for the current criterion but there is still room for improvement for the proposed tighter criterion.

Considering the underlying analytical performance, Table 1 shows that the bias was stable and low: 0.6 mmol/mol (0.05 %) in 2024 and 0.4 mmol/mol (0.04 %) in 2016–2024. The sustained small biases across nine years evidenced the efficacy of the IFCC global standardization program and demonstrates the power of a unified approach to analytical quality. The BLCV was consistent: 4.0 % (2.7 %) in 2024 and 4.3 % (2.9 %) in 2016–2024. This is strikingly comparable to the BLCV seen in recent CAP surveys in the USA (e.g. in survey GH5-C in 2024; mean BLCV in five specimens was 4.1 % (2.8 %)) [11]. Although the general picture suggests that performance has remained the same over the last nine years, this summary level overview masks significant differences in performance at a sub level e.g. by manufacturer groups.

From Supplementary Table 8 it can be seen that the application/use of analytical principles (HPLC, immunochemistry etc.) has remained remarkably consistent with small to moderate shifts of the market share of manufacturers.

## Performance by country (EQA organizer)

Table 3 shows that performance differed by country. Data from 2024 show that in the best performing countries more than 99 % of the laboratories met the current IFCC criterion and 92 % the proposed criterion. For the worst performing countries, these percentages were 90 and 65 % respectively. At a country level: the performance goal of 95 % of the laboratories within the criterion was met by 13 of the 16 countries for the current, and 0 of the 16 countries for the proposed criterion. This is visualized in more detail in Figure 2. Figure 2A shows that the bias was low for all countries but that the BLCV is substantially higher and variable from 3 to 6 % (2.0–4.1 %). The picture of 2024 (Panel D) is very similar: the same pattern of low bias and higher BLCV and the same countries are just outside the triangle in both panels. The general picture is that of persistent analytical performance from which the BLCV contributes more to total error (and thus to meeting or failing the criteria) than the bias.

## Performance by manufacturer

Table 4 shows that the range of performances across manufacturers is wider than those per country with extremes at the good and the poor ends. Data from 2024 show that with the best-performing methods >99 % of the laboratories met the current criterion and 99 % the proposed criterion. For the worst performing manufacturers these percentages are 70 and 42 %, respectively. At a manufacturer level: the performance goal of 95 % of users meeting the criteria was achieved by 21 of 25 manufacturers for the current, and 7 of 25 manufacturers for the proposed criterion. So, contrary to the countries where none met the proposed criterion, 7 manufacturers succeeded in achieving this goal. Mean results from 2016 to 2024 mirrored those of the 2024 data. This is visualized in more detail in Figure 3. Figure 3A shows the mean analytical performance over the years 2016–2024 for the major manufacturers. Colors relate the methods to their analytical principle. It can be seen that the performance was not strictly related to analytical principle. For ion exchange HPLC (red) and immunochemistry (green), there were good and poor performers. Manufacturers using enzymatic

**Table 3:** Performance per country (EQA Organizer) in analytical terms (SI units)[a] and failure rates of laboratories according to the current and proposed IFCC criteria.

| EQA organizer (country code) | 2016–2024 | | | | | 2024 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Analytical performance | | | % Laboratories failing | | Analytical performance | | | % Laboratories failing | |
| | n | Bias mmol/mol | Between laboratory CV | Current criterion 5 mmol/mol | Proposed criterion 3 mmol/mol | n | Bias mmol/mol | Between laboratory CV | Current criterion 5 mmol/mol | Proposed criterion 3 mmol/mol |
| Sciensano Belgium (BE) | 125 | +0.6 | 3.2 % | <1 % | 8 % | 116 | +0.7 | 3.1 % | <1 % | 8 % |
| SEKK s.r.o. Czech Republik (CZ) | 177 | +0.9 | 4.0 % | 2 % | 16 % | 184 | +1.1 | 4.1 % | 3 % | 20 % |
| Labquality Finland (FI) | 230 | −0.1 | 4.3 % | 2 % | 16 % | 198 | 0.0 | 4.1 % | 1 % | 15 % |
| CTCB France (FR1) | 136 | +0.3 | 3.9 % | 1 % | 13 % | 138 | +0.1 | 3.8 % | 1 % | 12 % |
| Asqualab France (FR3) | | | | | | 19 | +0.4 | 4.4 % | 3 % | 19 % |
| INSTAND Germany (DE1) | 628 | +0.3 | 4.5 % | 3 % | 19 % | 582 | +0.5 | 4.0 % | 2 % | 15 % |
| RfB Germany (DE2) | 670 | +0.6 | 4.2 % | 2 % | 19 % | 724 | +0.6 | 3.5 % | 1 % | 10 % |
| QualiCont Hungary (HU) | 70 | +0.6 | 5.6 % | 8 % | 29 % | 67 | +0.9 | 5.0 % | 6 % | 26 % |
| IEQAS Ireland (IE) | 45 | +0.5 | 3.7 % | 1 % | 11 % | 60 | +0.2 | 3.8 % | 1 % | 12 % |
| CRB Italy (IT1) | 52 | +0.9 | 4.7 % | 4 % | 24 % | 38 | +1.7 | 4.5 % | 7 % | 30 % |
| CRRVEQ Italy (IT2) | 112 | +0.9 | 4.9 % | 6 % | 26 % | 46 | +1.7 | 5.1 % | 10 % | 35 % |
| SKML Netherlands (NL) | 129 | +0.5 | 3.5 % | <1 % | 10 % | 129 | +0.5 | 3.4 % | <1 % | 9 % |
| SEQCML (now SEMEDLAB) Spain (ES) | 119 | +0.8 | 3.4 % | <1 % | 11 % | 149 | +1.2 | 4.1 % | 4 % | 21 % |
| Equalis Sweden (SE) | 115 | −0.1 | 3.8 % | 1 % | 13 % | 92 | +0.2 | 4.3 % | 3 % | 17 % |
| CSCQ Switzerland (CH) | 75 | 0.0 | 4.4 % | 2 % | 14 % | 54 | +0.4 | 4.4 % | 3 % | 18 % |
| Tübitak UME Turkey (TR1) | 48 | +0.7 | 6.2 % | 12 % | 35 % | | | | | |
| Weqas United Kingdom (UK) | 157 | +0.7 | 3.9 % | 2 % | 15 % | 157 | +0.9 | 4.0 % | 2 % | 17 % |
| Overall | 2,571 | +0.4 | 4.3 % | 3 % | 18 % | 2,753 | +0.6 | 4.0 % | 2 % | 16 % |

[a]Data in NGSP units are in Supplementary Table 10.

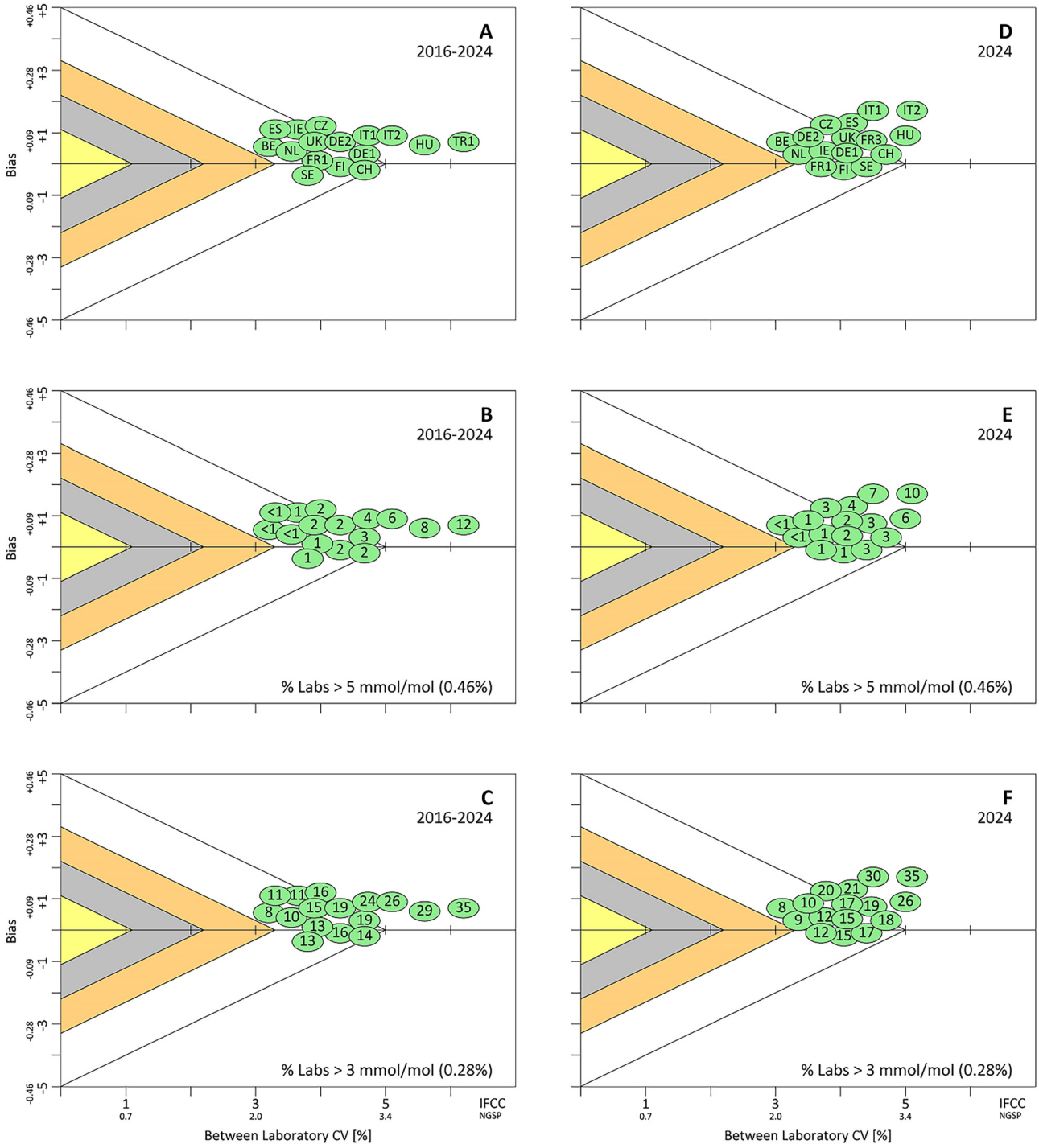


**Figure 2:** Results per country: Mean performance 2016–2024 and 2024 only. IFCC-MQT: BLCV (in both SI – and NGSP units) is on the horizontal axis and bias on the vertical axis. The white triangle represents the current performance criterion and amber, grey and yellow triangles challenging quality goals. Panels A, B, and C show the mean performance for 2016–2024; panels D, E, and F for 2024 only. Bias and BLCV are plotted in panels A and D. Percentages laboratories not meeting the current criterion of 5 mmol/mol (0.46 %) are in panels B and E. Percentages laboratories not meeting the proposed criterion of 3 mmol/mol (0.28 %) are in panels C and F. The ovals represent the respective countries: Belgium Sciensano (BE), Czech Republic SEKK s.r.o (CZ), Finland labquality (FI), France CTCB (FR1), France Asqualab (FR3), Germany INSTAND (DE1), Germany SPMD-RfB (DE2), Hungary QualiCont (HU), Ireland IEQAS (IE), Italy CRB (IT1), Italy CRRVEQ (IT2), Netherlands SKML (NL), Spain SEQC$^{ML}$ (now SEMEDLAB) (ES), Sweden Equalis (SE), Switzerland CSCQ (CH), United Kingdom Weqas (UK).

**Table 4:** Performance per manufacturer in analytical terms (SI units)[a] and failure rates of laboratories according to the current and proposed IFCC criteria.

| EQA organizer (manufacturer code) | 2016–2024 | | | | | 2024 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Analytical performance | | | % Laboratories failing | | Analytical performance | | | % Laboratories failing | |
| | n | Bias mmol/mol | Between laboratory CV | Current criterion 5 mmol/mol | Proposed criterion 3 mmol/mol | n | Bias mmol/mol | Between laboratory CV | Current criterion 5 mmol/mol | Proposed criterion 3 mmol/mol |
| Abbott Alinity (A) | 30 | −0.7 | 2.3 % | <1 % | 2 % | 43 | −1.3 | 2.8 % | <1 % | 12 % |
| Abbott Architect enzymatic (B) | 33 | −0.4 | 3.3 % | <1 % | 8 % | 38 | −1.3 | 2.0 % | <1 % | 5 % |
| Abbott/Alere Afinion (C) | 172 | −1.0 | 3.6 % | 2 % | 15 % | 262 | −0.5 | 3.7 % | 1 % | 12 % |
| ARKRAY HA 8160 (D) | 53 | +0.3 | 3.8 % | <1 % | 12 % | | | | | |
| ARKRAY HA 8180 (E) | 162 | +0.6 | 3.1 % | <1 % | 7 % | 179 | +0.3 | 2.8 % | <1 % | 3 % |
| ARKRAY HA 8190 (F) | 36 | +0.6 | 1.7 % | <1 % | <1 % | 44 | +0.8 | 1.9 % | <1 % | 1 % |
| Beckman Coulter AU series (G) | 52 | 0.0 | 5.7 % | 9 % | 30 % | 54 | +0.5 | 4.0 % | 2 % | 15 % |
| Bio-Rad D-10 (H) | 67 | +1.2 | 4.5 % | 5 % | 24 % | 81 | +0.4 | 3.9 % | 1 % | 14 % |
| Bio-Rad D-100 (I) | 67 | −0.6 | 2.8 % | <1 % | 5 % | 102 | −0.3 | 2.4 % | <1 % | 2 % |
| Bio-Rad Variant (J) | 96 | +0.8 | 4.7 % | 4 % | 23 % | 50 | +0.5 | 4.6 % | 3 % | 22 % |
| EKF Diagnostics (K) | 16 | +1.7 | 4.6 % | 7 % | 31 % | 17 | +2.4 | 3.1 % | 5 % | 35 % |
| HemoCue HbA$_{1c}$ 501 (L) | 21 | −1.3 | 7.4 % | 21 % | 46 % | 31 | −0.6 | 7.4 % | 19 % | 43 % |
| Menarini Hb next (M) | 26 | +2.2 | 4.4 % | 11 % | 38 % | 41 | +3.8 | 4.5 % | 30 % | 58 % |
| Roche cobas b 101 (N) | 19 | +0.1 | 3.6 % | <1 % | 11 % | 29 | −0.5 | 3.2 % | <1 % | 8 % |
| Roche cobas c 303/503 (O) | 97 | +1.5 | 3.1 % | 1 % | 17 % | 177 | +1.8 | 2.7 % | 1 % | 19 % |
| Roche cobas c 501/502 (P) | 248 | +0.8 | 3.5 % | 1 % | 12 % | 200 | +1.2 | 3.1 % | 1 % | 13 % |
| Roche cobas c 513 (Q) | 66 | +0.9 | 2.4 % | <1 % | 4 % | 76 | +1.1 | 2.1 % | <1 % | 4 % |
| Roche cobas Integra (R) | 55 | +0.8 | 3.5 % | 1 % | 12 % | 47 | +1.0 | 3.3 % | 1 % | 11 % |
| Sebia CAPILLARYS 2 (S) | 81 | −0.2 | 3.0 % | <1 % | 5 % | 31 | 0.0 | 2.4 % | <1 % | 1 % |
| Sebia CAPILLARYS 3 (T) | 97 | −0.1 | 2.5 % | <1 % | 1 % | 189 | −0.4 | 2.2 % | <1 % | 1 % |
| Siemens Atellica enzymatic (U) | 43 | −0.6 | 3.1 % | <1 % | 7 % | 51 | −1.1 | 3.0 % | <1 % | 11 % |
| Siemens DCA/Vantage (V) | 211 | +0.3 | 4.1 % | 1 % | 15 % | 179 | +0.3 | 5.3 % | 7 % | 26 % |
| Siemens Dimension (W) | 68 | +0.3 | 3.9 % | 1 % | 14 % | 27 | −1.1 | 2.8 % | <1 % | 9 % |
| Tosoh G8 (X) | 243 | +1.5 | 2.5 % | <1 % | 12 % | 146 | +1.7 | 2.6 % | <1 % | 16 % |
| Tosoh G11 (Y) | 173 | +1.3 | 2.3 % | <1 % | 7 % | 315 | +1.6 | 2.1 % | <1 % | 9 % |
| Trinity Biotech Premier Hb9210 (Z) | 28 | +1.3 | 3.8 % | 3 % | 20 % | 28 | +2.9 | 3.4 % | 11 % | 48 % |
| Overall | 2,571 | +0.4 | 4.3 % | 3 % | 18 % | 2,753 | +0.6 | 4.0 % | 2 % | 16 % |

[a]Data in NGSP units are in Supplementary Table 11.

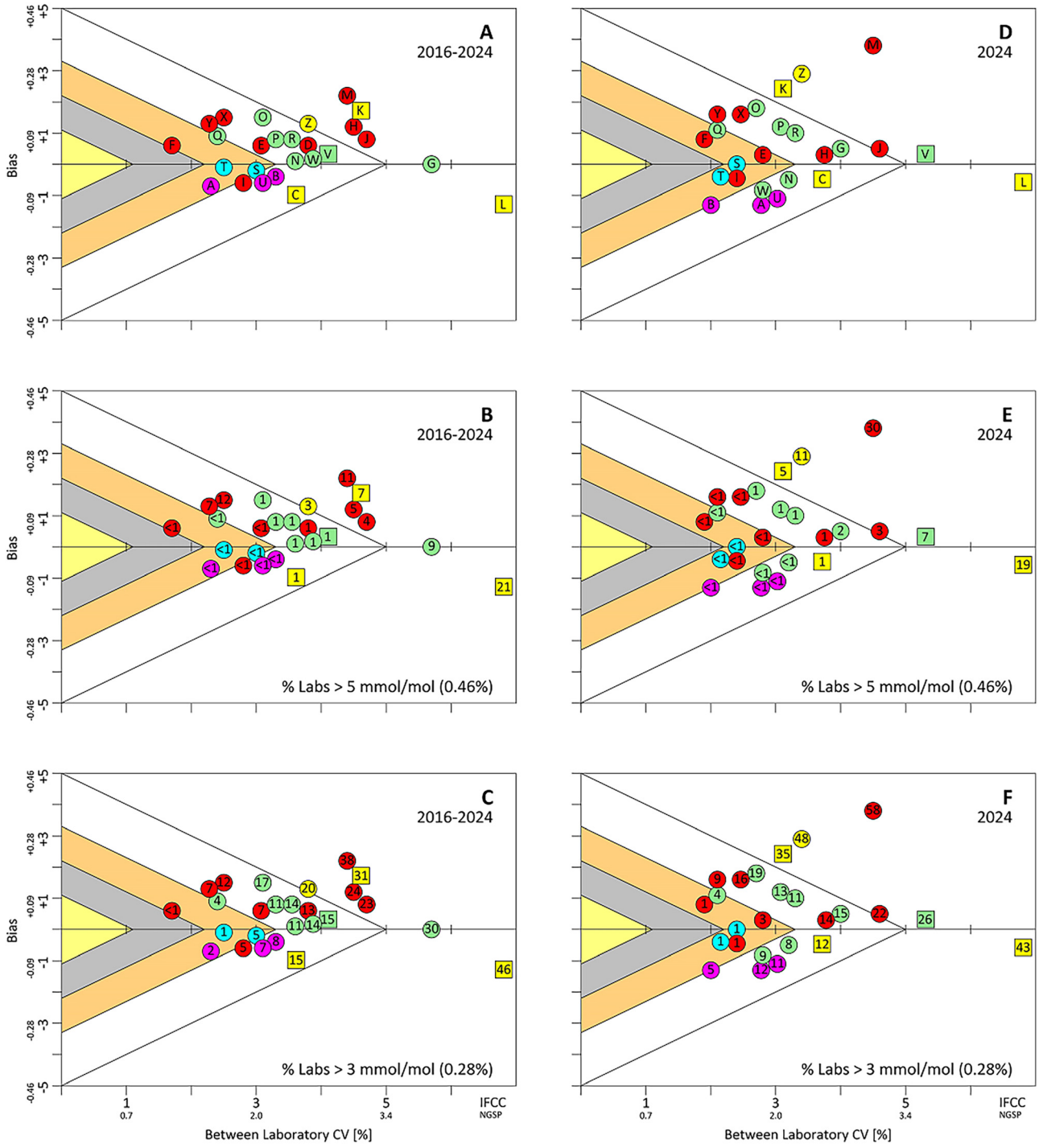


**Figure 3:** Results per manufacturer: Mean performance 2016–2024 and 2024 only. IFCC-MQT: BLCV (in both SI – and NGSP units) is on the horizontal axis and bias on the vertical axis. The white triangle represents the current performance criterion and amber, grey and yellow triangles challenging quality goals. Panels A, B, and C show the mean performance for 2016–2024; panels D, E, and F for 2024 only. Bias and BLCV are plotted in panels A and D. Percentages laboratories not meeting the current criterion of 5 mmol/mol (0.46 %) are in panels B and E. Percentages laboratories not meeting the proposed criterion of 3 mmol/mol (0.28 %) are in panels C and F. The dots (laboratory instruments) and squares (point-of-care instruments) show the manufacturers. Abbreviations: Abbott Alinity (A), Abbott Architect (B), Abbott Alere Afinion (C), ARKRAY HA-8160 (D), ARKRAY HA-8180 (E), ARKRAY HA-8190 (F), Beckman Coulter AU series (G), Bio-Rad D-10 (H), Bio-Rad D-100 (I), Bio-Rad Variant (J), EKF Diagnostics (K), HemoCue HbA$_{1c}$ 501 (L), Menarini Hb NEXT (M), Roche cobas b 101 (N), Roche cobas c 303/503 (O), Roche cobas c 501/502 (P), Roche cobas c 513 (Q), Roche Cobas Integra (R), Sebia CAPILLARYS 2 (S), Sebia CAPILLARYS 3 (T), Siemens Atellica (U), Siemens DCA/Vantage (V), Siemens Dimension (W), Tosoh G8 (X), Tosoh G11 (Y), Trinity Biotech Premier Hb9210 (Z).

**Table 5:** Non-commutability of lyophilized hemolysates per analytical Principle and per method, expressed as the matrix effect on bias and between-laboratory CV in SI units.

| Method principle | Method | Number results | | Bias mmol/mol | | | Between-laboratory CV% | | | Matrix effect | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WB | LH | WB | LH | Matrix effect (Exp. Unc.)[a] | WB | LH | Matrix effect | Bias | CV |
| Enzymatic | Abbott Alinity | 149 | 162 | −0.7 | −2.7 | −2.0 (0.5) | 2.3 | 5.8 | 2.7 | −1.9 | 2.7 |
| | Abbott Architect | 297 | 321 | −0.4 | −2.8 | −2.4 (0.4) | 3.3 | 6.7 | 3.4 | | |
| | Siemens Atellica | 130 | 85 | −0.6 | −2.0 | −1.4 (0.5) | 3.1 | 4.4 | 1.3 | | |
| Immuno chemistry | Beckman Coulter AU | 464 | 146 | 0.0 | 2.3 | 2.3 (0.7) | 5.7 | 7.3 | 1.6 | 1.7 | 1.4 |
| | Roche cobas c 501/502 | 1,487 | 935 | 0.8 | 2.2 | 1.4 (0.2) | 3.5 | 5.0 | 1.5 | | |
| | Roche cobas c 303/503 | 388 | 204 | 1.5 | 3.3 | 1.8 (0.3) | 3.1 | 4.1 | 1.0 | | |
| | Roche cobas c 513 | 397 | 103 | 0.9 | 2.6 | 1.7 (0.4) | 2.4 | 3.5 | 1.1 | | |
| | Roche cobas Integra | 332 | 179 | 0.8 | 1.9 | 1.1 (0.4) | 3.5 | 5.3 | 1.8 | | |
| | Siemens Dimension | 615 | 187 | 0.3 | 2.4 | 2.1 (0.4) | 3.9 | 5.3 | 1.4 | | |
| HPLC ion exchange | Bio-Rad D-10 | 607 | 614 | 1.2 | 0.5 | −0.7 (0.3) | 4.5 | 5.2 | 0.7 | −0.5 | 0.7 |
| | Bio-Rad D-100 | 607 | 498 | −0.6 | 0.0 | 0.6 (0.2) | 2.8 | 2.8 | 0.0 | | |
| | Bio-Rad Variant | 861 | 509 | 0.8 | 0.5 | −0.3 (0.3) | 4.7 | 6.2 | 1.5 | | |
| | ARKRAY HA-8160 | 425 | 375 | 0.3 | 0.2 | −0.1 (0.3) | 3.8 | 4.4 | 0.6 | | |
| | ARKRAY HA-8180 | 1,456 | 735 | 0.6 | −0.8 | −1.4 (0.2) | 3.1 | 4.5 | 1.4 | | |
| | Tosoh G7 | 155 | 193 | 1.1 | 1.5 | 0.4 (0.5) | 5.3 | 4.5 | 0.8 | | |
| | Tosoh G8 | 2,184 | 1,183 | 1.5 | 0.5 | −1.0 (0.1) | 2.5 | 3.7 | 1.2 | | |
| | Tosoh G11 | 1,214 | 712 | 1.3 | 0.8 | −0.5 (0.1) | 2.3 | 3.6 | 1.3 | | |
| | Tosoh GX | 87 | 111 | 1.6 | 0.1 | −1.5 (0.5) | 3.2 | 3.8 | 0.6 | | |
| HPLC affinity chrom | Trinity Premier Hb9210 | 353 | 155 | 1.3 | 0.4 | −0.9 (0.4) | 3.8 | 4.8 | 1.0 | −0.9 | 1.0 |
| Capillary Electrophoresis | Sebia CAPILLARYS 2 | 729 | 750 | −0.2 | −0.7 | −0.5 (0.2) | 3.0 | 3.3 | 0.3 | −0.4 | 0.1 |
| | Sebia CAPILLARYS 3 | 872 | 1,035 | −0.1 | −0.3 | −0.2 (0.1) | 2.5 | 2.5 | 0.0 | | |
| | Sebia MINICAP | 127 | 120 | −0.3 | −0.9 | −0.6 (0.4) | 3.5 | 3.5 | 0.0 | | |
| Point-of-care | Siemens DCA/Vantage | 1,901 | 287 | 0.3 | 5.2 | 4.9 (0.3) | 4.1 | 5.1 | 1.0 | 4.9 | |

[a]Expanded uncertainty.

(violet) or capillary electrophoresis (blue) methods tended to cluster together and performed well. Performance of point-of-care instruments (squares) was more dispersed, generally performing not as well as laboratory instruments (dots) but still better than some. It should be remarked that (an unknown) part of the POCT instruments were situated outside the laboratory which may have impacted the performance. Although data from 2024 (panel D) appears similar to that of 2016–2024 (panel A) closer inspection suggests that the BLCV has improved. This is confirmed by the data in Table 4: in 2024 50 % of the manufacturers had a BLCV below 3 % (2 % NGSP) whereas this was 25 % in 2016–2024. Overall, the analytical performance of the manufacturers was partly dependent on the analytical principle of the method and generally changed very little over the years. However, there were substantial differences between the manufacturers: some methods demonstrated excellent performance, however others fell significantly short of expected targets.

## Country vs. manufacturer performances

There is an interplay between the performance by country and by manufacturer. This can best be seen when Figures 2 and 3 are considered side by side. Results of countries tend to cluster along the horizontal axis (indicating a low bias) with a shift towards higher BLCVs whereas results of manufacturers show more bias but with, at least for many manufacturers, lower BLCVs. The performance of countries is the accumulation of performance of the laboratories in that country and the quality of the methods they used. If poor methods are used, the performance of the country will be poor, irrespective of the efforts of the laboratories. And even when acceptable methods are used, the performance might be fair: if in a country laboratories used methods X and Y (both with excellent BLCVs, but X having a positive and Y a negative bias), the BLCV of that country will be high due to the opposite effect of the biases of both methods used. This averaging out is a general phenomenon as can be seen in Figure 2: results of all countries had a low bias, but a BLCV that is higher than the BLCV of most manufacturers in Figure 3. This also explains why none of the countries met the proposed criterion while a number of manufacturers did.

## Commutability of lyophilized specimens

In the Eur$A_{1c}$ trial EQA organizers can choose whether they want to use WB or LH specimens. Lyophilized specimens

have the advantage of robust stability which facilitates easier logistics: no time-critical shipment, storage or planning issues. But the major disadvantage is the potential non-commutability of these specimens: measured bias might be due to non-commutability rather than to true bias. In addition, some methods (such as POCT instruments) cannot measure LH. WB specimens have no commutability issues, but the limited stability presents a logistical challenge. EQA organizers balance the pros and cons in their specific situation and make their own choice. As can be seen from Table 1 about two thirds of the specimens used were WB and one third LH.

Initial inspection of the overall results of the WB and LH specimens in Table 1 does not suggest non-commutability of LH: the mean bias over the years is +0.3 mmol/mol (0.3 %) vs. +0.4 mmol/mol (0.4 %) in WB. But the higher BLCV for LH (5.9 vs. 4.3 %; (4.0 vs. 2.9 % in NGSP units)) suggests non-commutability due to methods with a positive and methods with a negative effect of lyophilization on the bias. This translates to overall results indicating a bias comparable to WB (averaging out of results of methods with a positive and methods with a negative effect) but a higher BLCV.

Non-commutability is often described as the "matrix effect": the difference in bias between the LH and WB specimens. The IFCC Working Group on Commutability in Metrological Traceability has published several papers including guidance for evaluating the commutability of materials used in EQA [12]. Application of these models is complex and beyond the scope of this paper. Therefore a simplified approach was chosen rather than a full commutability assessment. A unique aspect of the EurA$_{1c}$ trial is that WB and LH specimens are prepared from the same blood which allows a direct comparison of results in both sample types, to investigate potential non-commutability. This allows us to address questions such as a) is there a matrix effect, b) is this effect relevant, c) is the matrix effect method-specific and/or related to the analytical principle, and d) is correction for the matrix effect possible?

Table 5 confirms the existence of a matrix effect, positive for some, negative for other methods. With +4.9 mmol/mol (0.45 %) this was most extreme for Siemens DCA/Vantage but with −2.0 and −2.4 mmol/mol (0.18 and 0.22 %) also substantial for the Abbott enzymatic methods. In some cases, the matrix effect appeared to correlate with the analytical principle: negative for enzymatic and positive for immunochemical methods. Ion exchange HPLC methods showed a diverse picture with both moderate positive and negative matrix effects. Capillary electrophoresis was minimally affected. Due to the large number of measurements, the expanded uncertainties, shown in Table 5, were low and statistically significant for nearly all methods. The critical question is whether these differences are also clinically relevant. There are no ISO standards for non-commutability but ISO 13528 states that non-homogeneity of EQA specimens should contribute <10 % to the BLCV. This criterion might also be applied to non-commutability: the matrix effect of EQA specimens should contribute <10 % to the BLCV. The $S_{PT}$ is BLCV x mean HbA$_{1c}$ concentration (mean 2016–2024) divided by 100 %=5.9×49.8/100=3 mmol/mol (0.03 %). Or in practical terms: $S_s<0.3\sigma S_{PT}$ in which $S_s$ is the SD of inhomogeneity and $S_{PT}$ the SD of BLCV. Then the maximum allowable contribution of the matrix effect is 0.3×3=0.9 mmol/mol (0.08 %). As can be seen from Table 5, this criterion was only met for most ion exchange HPLC and all capillary electrophoresis methods. In cases where the matrix effect does not meet the allowable criterion, method specific target values can be used. Alternatively, results could be corrected for the matrix effect. This is explored for the results of the LH specimens of the 2023 round. Results of the laboratories using the methods in Table 5 (thus the methods for which the matrix effect is known) were corrected for the matrix effect and overall BLCV was calculated from those corrected results. Example: From Table 5 it can be seen that the matrix effect for Abbott Architect method was −2.4 mmol/mol (0.22 %) and for the Beckman Coulter AU was +2.3 mmol/mol (0.21 %). Results of Abbott users were corrected by adding 2.4 mmol/mol (0.22 %) and those of Beckman Coulter users by subtracting 2.3 mmol/mol (0.21 %). In this way, the BLCV dropped from 5.6 to 4.4 % (3.8–3.0 %), which is quite close to the BLCV of 3.9 % (2.7 %) seen for the WB using laboratories. It can be concluded that there is a potential for correction for non-commutability by using method specific target values, at least for the type of LH used in the trial.

In summary: a) there was a matrix effect, b) this effect was potentially clinically relevant for a number of methods, c) the matrix effect was partly method and partly analytical principle related, and d) the matrix effect could potentially be corrected. However, it is still preferable to use WB specimens.

## Perfect is the enemy of good

The introduction states that EQA, by identifying poor performers (laboratories, countries and methods) stimulates quality improvement. But a major conclusion of the nine years of the EurA$_{1c}$ trial is that the quality has changed very little. This raises the question as to why? There have indeed been positive developments: Abbott and Siemens replaced their poor performing immunoassays with better performing enzymatic methods, Tosoh, ARKRAY and Bio-Rad introduced new, better performing generations, and Roche developed a dedicated instrument for HbA$_{1c}$. Unfortunately,

this improvement has been offset by the introduction of poor performing methods (M and L in Figure 3D) or ones which appear to have temporarily lost their traceability to the IFCC RMP (Z in Figure 3D). Looking forward, a critical question is whether it is possible to achieve better performance/quality in the future. It seems obvious that avoiding the use of poor methods would be a logical step. Perfect traceability of all methods has a marginal effect: when results of the 2024 trial were corrected for the method specific bias, the overall bias dropped to zero (per definition!) and the BLCV from 4.0 to 3.7 % (2.7–2.5 %). A focus on improvement of the BLCV should be a primary goal. Manufacturers should aim to reduce batch-to-batch variation through more stringent management of value assignment to kit calibrators and reagent productions. (National) regulatory requirements for analytical performance and laboratory quality might be another driver for quality improvement. Further, laboratory managers should focus on more rigorous quality management processes and enhanced education of laboratory staff. Joint efforts by manufacturers and their users may not only lead to the current IFCC criterion being met more fully in the future, but also the proposed tighter criterion, so that HbA$_{1c}$ can be used safely and reliably for both monitoring and diagnosis of diabetes.

## Limitation of the trial

A limitation of the EurA$_{1c}$ trial was the diversity of data submission and reporting systems of the EQA organizers. This made it complicated and time consuming to collate the data. In a few cases (approximately 2 % of the results) it was impossible to unambiguously identify the method used by the participating laboratories and these had to be classified as unknown. In future it would be helpful if EQA organizers would apply the guidelines of the European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM) [13]. This umbrella organization of EQA organizers advocates collaborative studies (such as the EurA$_{1c}$ trial) as a strong driver for quality improvement [14]. The focus of most EQA-organizers is on medical laboratories and therefore POCT systems (often outside laboratories) are under-represented in the EurA$_{1c}$ trial.

# Concluding statements

(1) Overall performance of HbA$_{1c}$ measurement has been stable over the past decade.
(2) The consistently low bias over the years indicates the continued success of IFCC standardization efforts.
(3) Breaking down the results by country for 2016–2024 shows that in the best-performing country, 100 % of laboratories met the current criterion and 92 % would meet the proposed tighter criterion. For the worst-performing countries, the percentages were 88 and 65 %, respectively.
(4) Differentiation by manufacturer reveals significant performance differences. Users of the best methods achieved the current criterion 100 % of the time and the proposed criterion 99 % of the time. For the worst-performing method, the percentages were 79 and 54 %, respectively.
(5) There is a causal relationship between performance by country and by manufacturer. Within countries, the results from methods with positive and negative biases are averaged out. Consequently, countries have relatively low bias and higher BLCV. The opposite is true for manufacturers.
(6) LH specimens, especially for enzymatic and immunochemical methods, have a small but significant degree of non-commutability. This matrix effect could potentially be corrected, but it would be preferable to use WB specimens in EQA programs.
(7) There is scope for improvement of performance. This can be achieved by avoiding the worst performing methods. For manufacturers this means focusing on more stringent batch-to-batch management and more robust methods and for laboratories paying more attention to quality management processes.
(8) The EurA$_{1c}$ trial benefits from a robust, large dataset across a range of countries and manufacturers. As such, it is a powerful tool for monitoring the success of IFCC standardization efforts, accurately identifying poorly performing methods, and thus pushing for quality improvement and enforcement.

**Acknowledgments:** The authors thank all laboratories for their participation in the EurA$_{1c}$ trial.
**Research ethics:** Not applicable.
**Informed consent:** Not applicable.
**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
**Use of Large Language Models, AI and Machine Learning Tools:** None declared.
**Conflict of interest:** The authors state no conflict of interest.
**Research funding:** None declared.
**Data availability:** Not applicable.
**Employment or leadership:** David Sacks: This research was supported (in part) by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy

requirements, and are considered works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH of the U.S. Department of Health and Human Services.

## EurA1c Trial Group

**IFCC C-EUBD**

Emma English, University of Cambridge, Cambridge, United Kingdom.
Kyunghoon Lee, Seoul National University Bundang Hospital, Seoul, South Korea.
Jan Skrha jr., Charles University in Prague, Prague, Czech Republic.
Eric Kilpatrick, Manchester University, Manchester, United Kingdom.
Liza Kunz, Abbott, Chicago, USA.
Erna Lenters-Westra, Isala and European Reference Laboratory, Zwolle, Netherlands.
Carla Siebelder, Queen Beatrix Hospital and European Reference Laboratory, Winterswijk, Netherlands.
David Sacks, National Institutes of Health, Bethesda, USA.
Annalise Zemlin, NHLS, Cape Town, South Africa.
Cas Weykamp, Queen Beatrix Hospital and European Reference Laboratory, Winterswijk, Netherlands.

**EQA Organizers**

Austria, ÖQUASTA, Vienna: Christoph Buchta.
Belgium, Sciensano, Brussels: Yolande Lenga.
Czech Republic, SEKK s.r.o., Pardubice: Marek Budina, Josef Kratochvila, Ondrej Wiewiorka.
Finland, Labquality, Helsinki: Päivi Ranta.
France, ASQUALAB, Paris: Anne Vassault, Ingrid Anglard.
France, CTCB, Toulouse: Safouane Hamdi, Stéphanie Albarède.
France, PROBIOQUAL, Lyon: Philippe Joly, Bernard Poggi.
Germany, INSTAND e.V., Düsseldorf: Patricia Kaiser.
Germany, Reference Institute for Bioanalytics, Bonn: Anja Kessler, Marika Enders, Caroline Stobe, Birgit Zinndorf.
Greece, ESEAP, Athens: Alexander Haliassos, Serafeim Karathanos, Lia Moira, Efi Botoula.
Hungary, QualiCont, Szeged: Erika Sárkány, Petra Magdolna Molnár.
International Group, ERL, Winterswijk: Carla Siebelder, Erna Lenters.
Ireland, IEQAS, Dublin: Anne Kane, Hazel Graham, Marguerite MacMahon, Phyllis Reilly.
Italy, Centro Ricerca Biomedica per la Qualità in Medicina di Laboratoria, Padova: Laura Sciacovelli, Mario Plebani.
Italy, CRRVEQ, Florence: Paola Pezzati, Sabrina Buoro.
Korea, KEQAS, Seoul: Junghan Song.
Mexico, Grupo ERL, Panuco: Eduardo Rojano.
Netherlands, SKML, Nijmegen: Hans van der Vuurst, Marc Thelen.
Portugal, PNAEQ-INSA, Lisbon: Ana Faria, Helena Correia, Ana Cardoso.
South Africa, NHLS, Cape Town: Division of Chemical Pathology, Stellenbosch University and National Health Laboratory Service (NHLS), Annalise Zemlin.
Spain, SEMEDLAB, Barcelona: Montserrat Ventura Alemany, Berta Piqueras Andrés, Mariona Panadès Turró, Sandra Bullich Marin.
Sweden, Equalis AB, Uppsala: Jim Andersson, Carolina Kristoffersson.
Switzerland, CSCQ, Chêne-Bourg: Pierre-Alain Morandi, Dagmar Kesseler.
Thailand, NIH, Nonthaburi: Supaporn Suparak.
Turkey, TÜBİTAK UME, Gebze/Kocaeli: Gonca Altin, Diler Aslan, Müslüm Akgöz, Fatma Akcadag.
Turkey, KBUDEK, Istanbul: E. Cuneyt Canbulat.
UK, Weqas, Cardiff: Mary Annette Thomas, Gareth John Davies, Samantha Jones.
Vietnam, Quality Control Center for Medical Laboratory, Hanoi: Le Vu Huyen Trang, Pham Thi Huong Trang, Nguyen Tran Phuong, Trinh Thi Phuong Dung, Tran Huy Thinh.

**Reference laboratories**

France, University Hospital Center of Reims, Reims: Stéphane Jaisson, Philippe Gillery.
Germany, INSTAND e.V., Düsseldorf: Patricia Kaiser.
Germany, Reference Institute for Bioanalytics, Cologne: Caroline Stobe, Birgit Zinndorf.
Italy, Università degli Studi di Milano, Milano: Renata Paleari.
Netherlands, Isala, Zwolle: Erna Lenters, Robbert Slingerland, Janine Slootstra.
Netherlands, Queen Beatrix Hospital, Winterswijk: Carla Siebelder, Sanne Leppink, Laura Reijnders, Rogier Schoeman.

**Executive team**

Carla Siebelder, Cas Weykamp, Liesbeth Janssen, Marieke te Winkel, Irene de Graaf: Winterswijk, the Netherlands. Erna Lenters-Westra, Zwolle, the Netherlands.

## References

1. American Diabetes Association. Standards of care in diabetes – 2025. Diabetes Care 2024;48:S27–49.
2. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. Approved IFCC reference method for the measurement of HbA1c in human blood. Clin Chem Lab Med 2002;40:78–89.

3. Weykamp C, John WG, Mosca A, Hoshino T, Little R, Jeppsson JO, et al. The IFCC reference measurement system for HbA1c: a 6-year progress report. Clin Chem 2008;54:240–8.
4. Weykamp C, John G, Gillery P, English E, Ji L, Lenters-Westra E, et al. IFCC task force on implementation of HbA1c standardization. Investigation of 2 models to set and evaluate quality targets for HbA1c: biological variation and sigma-metrics. Clin Chem 2015;61:752–9.
5. Biological Variation Database of the European Federation of Clinical Chemistry and Laboratory Medicine. https://biologicalvariation.eu (Accessed 4 January 2026).
6. The EurA1c Trial Group. EurA1c: the European HbA1c trial to investigate the performance of HbA1c assays in 2166 laboratories across 17 countries and 24 manufacturers by use of the IFCC model for quality targets. Clin Chem 2018;64:1183–92.
7. Consensus Committee. Consensus statement of the worldwide standardization of hemoglobin A1c measurement: the American diabetes organization, European Association for the Study of Diabetes, International Federation of Clinical Chemistry and Laboratory Medicine, and the International Diabetes Federation. Diabetes Care 2007;30:399–400.
8. Weykamp CW, Mosca A, Gillery P, Panteghini M. The analytical goals for hemoglobin A1c measurement in IFCC units and National Glycohemoglobin Standardization Program units are different. Clin Chem 2011;57:1204–5.
9. Annual reports of EurA1c. https://www.ifcchba1c.org/GeneralInformation.aspx?I= (Accessed 20 January 2026).
10. Weykamp C, Siebelder C, Lenters E, Slingerland R, English E. The risk of clinical misinterpretation of HbA1c: modelling the impact of biological variation and analytical performance on HbA1c used for diagnosis and monitoring of diabetes. Clin Chim Acta 2023;548:117495.
11. College of American Pathologists GH5 Survey c2024 Summary Report. https://ngsp.org/CAP/CAP24c.pdf (Accessed 20 January 2026).
12. Sandberg S, Fauskanger P, Johansen J, Keller T, Budd J, Greenberg N, et al. Recommendations for setting a criterion for assessing commutability of sample materials used in external quality assessment/proficiency testing schemes. Clin Chem 2023;69:1227–37.
13. Buchta C, Gidske G, Henriksen GM, Badrick T. European Organisation of external quality assurance providers in laboratory medicine (EQALM). The European Organisation of external quality assurance providers in laboratory medicine (EQALM) statement: guidelines for publishing about interlaboratory comparison studies (PubILC). Crit Rev Clin Lab Sci 2024;61:588–98.
14. Buchta C, Aberle SW, Albarede S, Albe X, Badrick T, Bietenbeck A, et al. The concept of external quality assessment super challenges with special consideration of their importance during pandemics. J.lanmic 2025:101292. https://doi.org/10.1016/j.lanmic.2025.101292.

**Supplementary Material:** This article contains supplementary material (https://doi.org/10.1515/cclm-2026-0276).