

Gro Gidske*, Sverre Sandberg, Pernille Fauskanger, Jonna Pelanti, Mette C. Tollånes, Anne E. Solsvik, Una Ø. Sølvik, Wenche S. Vie and Anne Stavelin

Aggregated data from the same laboratories participating in two glucose external quality assessment schemes show that commutability and transfers of values to control materials are decisive for the biases found

<https://doi.org/10.1515/cclm-2023-0532>

Received May 20, 2023; accepted July 4, 2023;

published online July 21, 2023

Abstract

Objectives: We report the results of glucose measurements performed during one year by the same measurement procedures (MPs) in 58 Norwegian hospital laboratories using control materials provided by external quality assessment (EQA) schemes from two different providers. The providers used materials with presumed vs. verified commutability and transfers of values using reference material vs. using a highest-order reference MP.

Methods: Data from six Labquality and three Noklus glucose EQA surveys were aggregated for each MP (Abbott

Alinity, Abbott Architect, Roche Cobas, and Siemens Advia) in each scheme. For each EQA result, percent difference from target value (% bias) was calculated. Median percent bias for each MP per scheme was then calculated.

Results: The median % biases observed for each MP in the Labquality scheme were significantly larger than those in the Noklus scheme, which uses verified commutable control materials and highest-order reference MP target values. The difference ranged from 1.2 (Roche Cobas, 2.9 vs. 1.7%) to 4.4 percentage points (Siemens Advia, 3.2% vs. -1.2%). The order of bias size for the various MPs was different in the two schemes. In contrast to the Labquality scheme, the median % biases observed in the Noklus scheme for Abbott Alinity (-0.1%), Abbott Architect (-0.5%), and Siemens Advia (-1.2%) were not significantly different from target value ($p > 0.756$).

Conclusions: This study underlines the importance of using verified commutable EQA materials and target values traceable to reference MPs in EQA schemes designed for assessment of metrological traceability of laboratory results.

Keywords: commutability; data aggregation; external quality assessment; glucose; metrological traceability; systematic deviation

***Corresponding author: Gro Gidske**, MSc, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, P.O. Box 6165, N-5892 Bergen, Norway; and Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway, Phone: +47 90098912, E-mail: gro.gidske@noklus.no

Sverre Sandberg, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway; and Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway

Pernille Fauskanger, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Mathematics, University of Bergen, Bergen, Norway

Jonna Pelanti, Labquality Ltd, Helsinki, Finland

Mette C. Tollånes, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

Anne E. Solsvik, Wenche S. Vie and Anne Stavelin, The Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway

Una Ø. Sølvik, Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

Introduction

Glucose measurements are essential for the diagnosis and monitoring of diabetes mellitus. Since international diagnostic decision limits [1, 2] are independent of the measurement procedure (MP), glucose measurements should be accurate and equivalent regardless of which laboratory or MP used [3]. The performance of glucose measurements can be monitored by external quality assessment (EQA). Ideally, an EQA scheme should circulate commutable materials that are measured in replicates by the participating laboratories to assess the trueness of their MP compared to a reference MP, or a certified reference material [4]. However, the

different EQA schemes vary in their ability to provide this trueness verification, and one of the most important causes is lack of commutable materials [3].

Commutability is an important property of an EQA material (EQAM) that exhibits equivalent relative responses for the EQAM and the intended clinical samples, in method comparison studies involving two or more MPs for the same measurand [5]. Commutable EQAMs are rare because they are challenging to produce and distribute. EQA schemes with non-commutable EQAMs often exhibit greater stability and offer a wider range of concentrations and measurands. Moreover, EQAMs are often not assessed for commutability when fresh or fresh-frozen human materials with no additives are used, as their properties are assumed equivalent to those of clinical samples [3]. However, pooling, freezing or other storage conditions may change the properties of the EQAMs, so that they are no longer commutable with clinical samples [6]. When non-commutable EQAMs are used, reference target values cannot be used, and the EQA result can only be used to compare the analytical performance of an individual laboratory to that of other laboratories using the same MP. Additionally, if non-commutable certified reference materials are used in target value determination, the lack of full traceability to a highest-order reference MP can lead to erroneous conclusions regarding measurement trueness [3].

Hospital laboratories in Norway participate in a general chemistry scheme from the EQA provider Labquality (Helsinki, Finland) where glucose is one of 51 measurands. However, a positive bias in glucose measurements across various MPs within this scheme was observed some years ago. Therefore, since May 2021, Norwegian hospital laboratories have also participated in an EQA scheme designed and optimized for trueness assessment of glucose measurements, provided by the Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus) (Bergen, Norway). Labquality uses presumed commutable EQAMs with target values established by a certified reference material. Noklus uses EQAMs with verified commutability and target values established by a reference MP. To our knowledge, there are no studies that describe EQA results obtained over time by the same laboratories participating in two different EQA schemes from different providers. The aim of the present study was to compare the EQA results obtained by the same laboratories participating in the Labquality and the Noklus EQA schemes for serum glucose.

Materials and methods

Individual glucose EQA results for 58 Norwegian hospital laboratories participating in Labquality (Helsinki, Finland) and Noklus (Bergen,

Norway) EQA schemes were collected. The one-year data collection period (May 2021 to June 2022) included six Labquality surveys and three Noklus surveys, with two EQAMs in each survey.

Labquality EQA scheme “2050 serum B and C for general chemistry”

Preparation of minimally processed EQA materials: Labquality had the minimally processed EQAMs (serum B) prepared by one manufacturer (manufacturer I) by a process designed to ensure sample commutability [7]. One Fenwal blood bag without anticoagulant (Fenwal Laboratories, Deerfield, IL, USA) per donor was drawn from each blood donor. The blood was allowed to clot at room temperature, then centrifuged and pooled. The pools were stored frozen at -80°C until thawing, filtering, mixing, and dispensing into 3 mL vials and stored at -80°C . Each pool was assessed for bacterial contamination and found to be negative. Homogeneity was tested and approved in accordance with ISO 13528 [8].

Preparation of processed EQA materials: The processed EQAMs (serum C) were commercial fresh frozen EQAMs prepared by three different manufacturers (manufacturers II–IV) by their processes. The materials were of human blood origin (Table 1). Homogeneity was either tested and approved by the manufacturer or by Labquality in accordance with ISO 13528 [8].

Commutability assessment of EQA materials: The Labquality EQAMs have not been formally assessed for commutability.

Distribution and sample analysis: Labquality offers six surveys each year in the EQA scheme for general chemistry. In each survey, two EQAMs with varying glucose concentrations (Table 1) were distributed at ambient temperature. If laboratories were not able to analyse the EQAMs on the day of arrival, they were instructed to store them refrigerated until analysis. Glucose is one of 51 potential measurands to be measured. Based on results from the participating laboratories, the EQAMs were stable up to 22 days after distribution. The laboratories were instructed to measure serum glucose in duplicate and report mean values.

Target values: The target value is a transferred value from the Nordic Federation of Clinical Chemistry (NFKK) Reference Serum X (RSX), which is an unmodified fresh frozen human serum having certified values for several measurands, including glucose. The certified glucose value in RSX is traceable to the International Measurement Evaluation Program (IMEP)-17 material [9], whose target values were assigned by a reference MP using IDMS [10]. Five Nordic laboratories, using different MPs, measure serum B, serum C and RSX in triplicates in each survey. The transferred values (T) for the two EQAMs are then calculated as: $T = \frac{\text{mean of EQA sample} \times (\text{certified value for RSX})}{\text{mean of RSX}}$. Further calculations are made on these T -values after testing for outliers with Dixon’s Q-test [11]. The mean of the transferred values from the five laboratories is used as the target value. The standard uncertainty of the target value is calculated as standard error of the mean (SEM) of the T -values. The uncertainty of the certified value for RSX is included in the uncertainties shown in Table 1. Acceptance limits in the Labquality scheme are target value $\pm 6\%$.

Table 1: Overview of EQA materials used in Labquality and Noklus surveys from May 2021 to June 2022, number of laboratories and response rates.

Scheme	Survey	Sample characteristics					Participants (n)										Response rate, %
		Sample	Glucose target value, mmol/L ^d	EQA material	Manufacturer	Communitability assessment	Abbott Alinity	Abbott Architect	Roche Cobas	Siemens Advia	Other	No answer	Total				
Labquality	Survey 3, 2021	Serum B	5.43 ± 0.05	Frozen human serum, min. processed	I	Not formally assessed	7	6	27	4	2	12	58	79			
		Serum C ^b	4.23 ± 0.04	Frozen human serum, processed	II	Not formally assessed	6	6	29	4	2	11	58	81			
	Survey 4, 2021	Serum B	5.76 ± 0.06	Frozen human serum, min. processed	I	Not formally assessed	9	6	32	5	4	2	58	97			
		Serum C ^b	4.24 ± 0.05	Frozen human serum, processed	II	Not formally assessed	10	6	32	5	4	1	58	98			
	Survey 5, 2021	Serum B	3.92 ± 0.04	Frozen human serum, min. processed	I	Not formally assessed	8	4	22	4	0	20	58	66			
		Serum C	14.36 ± 0.20	Frozen human serum, processed	III	Not formally assessed	8	4	22	4	0	20	58	66			
	Survey 6, 2021	Serum B	5.49 ± 0.06	Frozen human serum, min. processed	I	Not formally assessed	8	5	31	4	3	7	58	88			
		Serum C ^a	3.95 ± 0.06	Frozen human serum, processed	II	Not formally assessed	8	5	31	4	3	7	58	88			
	Survey 1, 2022	Serum B	2.86 ± 0.05	Frozen human serum, min. processed	I	Not formally assessed	8	4	32	4	5	5	58	91			
		Serum C	14.45 ± 0.26	Frozen human serum, processed	IV	Not formally assessed	8	4	32	4	5	5	58	91			
Noklus	Survey 3, 2022	Serum B	5.44 ± 0.07	Frozen human serum, min. processed	I	Not formally assessed	9	4	31	3	4	7	58	88			
		Serum C ^a	4.01 ± 0.05	Frozen human serum, processed	II	Not formally assessed	9	4	31	3	4	7	58	88			
	Survey 1, 2021	Noklus ^c	5.52 ± 0.06	Frozen human serum, min. processed	Noklus	Previously batch formally assessed in 2020	11	5	32	5	4	1	58	98			
		Noklus	7.13 ± 0.07	Frozen human serum, min. processed	Noklus	Formally assessed in accordance with CLSI EP14-A3	11	5	32	5	4	1	58	98			
	Survey 2, 2021	Noklus ^c	5.52 ± 0.06	Frozen human serum, min. processed	Noklus	Previously batch formally assessed in 2020	12	4	32	4	5	1	58	98			
	Noklus	11.1 ± 0.1	Frozen human serum, min. processed	Noklus	Formally assessed in accordance with CLSI EP14-A3	12	4	32	4	5	1	58	98				

Table 1: (continued)

Scheme	Survey	Sample characteristics				Participants (n)							Response rate, %		
		Sample	Glucose target value, mmol/L ^d	EQA material	Manufacturer	Commutability assessment	Abbott Alinity	Abbott Architect	Roche Cobas	Siemens Advia	Other	No answer		Total	
	Survey 1, 2022	Noklus	3.52 ± 0.04	Frozen human serum, min. processed	Noklus	Previous batch formally assessed in 2020	10	3	25	4	4	4	12	58	79
		Noklus	5.42 ± 0.05	Frozen human serum, min. processed	Noklus	Formally assessed in accordance with CLSI EP14-A3	10	3	25	4	4	4	12	58	79

^{a, b}: Same EQA materials used in two EQA surveys. ^d: Lab quality target values: the uncertainty of the certified value for reference serum X (RSX) is included in the expanded uncertainties. Noklus target values: the uncertainty of the reference measurement procedure value is the expanded uncertainty ($k=2$). EQA, external quality assessment; min, minimally.

Noklus EQA scheme “serum glucose”

Preparation of EQA materials: The glucose EQAMs were prepared by Noklus. One Fenwal blood bag without anticoagulant (Fenwal Laboratories, Deerfield, IL, USA) was drawn from each blood donor at the Haukeland University Hospital blood bank, Bergen, Norway. The blood was allowed to clot at room temperature, then centrifuged at 3,500 revolutions per minute (RPM) for 12 min, and serum transferred to a new blood bag without anticoagulant before undergoing a second round of centrifugation. Serum from several donors were pooled. Some pools were kept without additives, whereas others had D(+)-glucose monohydrate dissolved in sodium chloride (Merck KGaA, Darmstadt, Germany) added to achieve various glucose concentrations (Table 1). The EQAMs were dispensed into 2 mL vials and stored at -80°C . Each pool was assessed for bacterial contamination and found to be negative. Homogeneity was tested and approved in accordance with ISO 13528 [8].

Commutability assessment of EQA materials: Noklus assessed the commutability of the EQAMs in 2020. The study was conducted in accordance with CLSI EP14-A3 [5]. In total, 25 patient serum samples from individuals with and without diabetes mellitus were drawn on 8.5 mL BD Vacutainer[®] SST[™] II serum tubes (BD Vacutainer Systems, Plymouth, UK). The patient serum glucose concentrations ranged from 4.2 to 13.2 mmol/L. Ideally, the patient samples used in a commutability assessment should consist of fresh samples [5]. However, obtaining fresh samples in the broad concentration range needed was not possible. Still, previous studies have shown that glucose is stable through freeze-thaw cycles [12, 13]. The three control batches included had glucose added to the concentrations of 5.42 ± 0.05 , 7.13 ± 0.07 and 11.1 ± 0.1 mmol/L (Table 1). The four Norwegian hospital MPs included were Abbott Alinity c (Abbott Diagnostics, IL, USA), Ortho Vitros 5,1FS (Ortho Clinical Diagnostics, NY, USA), Roche Cobas 6000 (Roche Diagnostics, Rotkreuz, Switzerland) and Siemens Advia Chemistry XPT (Siemens Healthcare Diagnostics, Erlangen, Germany).

Both patient samples and EQAMs were shipped on dry ice to the four laboratories. All samples were measured in triplicate during one working day, the three EQAMs interspersed among the 25 patient samples. Internal quality control was measured before and after the samples to verify MP stability. The EQAMs were assessed as commutable for all instrument combinations in all three glucose concentrations because their values were included in the 95 % prediction intervals calculated from the Deming regression of the 25 patient samples [5].

Distribution and sample analysis: Noklus offers two surveys each year in the EQA scheme for serum glucose. In each survey, two EQAMs with varying glucose concentrations (Table 1) were distributed at ambient temperature. If laboratories were not able to analyse the EQAMs on the day of arrival, they were instructed to store them refrigerated until analysis. Glucose is the only measurand to be measured. Based on stability testing in accordance with ISO 13528 [8], the EQAMs were stable up to ten days after distribution. The laboratories were instructed to measure serum glucose in five replicates and report mean values.

Target values: Target values were assigned by a reference MP using gas chromatography-isotope dilution mass spectrometry (GC-IDMS) by INSTAND e.V. (Düsseldorf, Germany), which is included in the Joint Committee for Traceability in Laboratory Medicine (JCTLM) database [14]. EQAMs were shipped on dry ice to the laboratory. Each of the EQAMs were analysed in duplicate on three different days, i.e., six

analyses in total, and the mean was used as the target value. Target values were assigned once for each batch (Table 1). Acceptance limits in the Noklus scheme are target interval $\pm 5\%$ (target interval=target value ± 0.1 mmol/L).

Reliability and reproducibility of the target values

Both Labquality and Noklus have conducted analyses to assess the reliability and reproducibility of the target values. For Labquality, the transferred target value of serum B in the third survey in 2022 (LQ3-22) was 5.44 ± 0.07 mmol/L (Table 1). Another sample of the same batch was sent to INSTAND e.V. for analysis with a reference MP, and the value 5.37 ± 0.05 mmol/L was comparable to the transferred target value. Also, two of the EQAMs used during this one-year period were used in two separate surveys having comparable transferred target values: Serum C in EQ3-21 (4.23 ± 0.04 mmol/L) and EQ4-21 (4.24 ± 0.05 mmol/L), and serum C in LQ6-21 (3.95 ± 0.06 mmol/L) and LQ3-22 (4.01 ± 0.05 mmol/L) (Table 1). In addition, Labquality sent a sample of RSX to INSTAND e.V. for analysis, and the result indicates that RSX is stable (4.405 ± 0.034 mmol/L in 2002 and 4.37 ± 0.04 mmol/L in 2021).

Noklus sent a sample that had been stored for nine years in -80°C , to INSTAND e.V. for analysis, and the result indicates that the EQAMs are stable and that the target value is reliable (5.71 ± 0.01 mmol/L by Ghent University in 2012 and 5.72 ± 0.06 mmol/L by INSTAND e.V. in 2021).

Data analysis

All statistical analyses were performed using the data.table, openxlsx, readxl and stringi packages in R software version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) and SPSS 28.0 software (SPSS Inc., IL, USA).

In cases where laboratories had two subscriptions (two instruments with same MP) in either the Labquality scheme ($n=11$) or the Noklus scheme ($n=1$) and reported two sets of results in an EQA survey, the mean of the two results was used in further calculations.

For each result, percent difference from target value (% bias) was calculated according to the formula: $\% \text{ bias} = 100 \times [(\text{measured concentration by the laboratory}) - (\text{target value})] / (\text{target value})$, using all decimals reported (up to three in Noklus scheme and unlimited in Labquality scheme). All % bias calculations were rounded to one decimal place.

MPs used by four or more participating laboratories were included in the study. Shapiro-Wilk test of normality showed that neither the data from the Labquality surveys ($n=556$) nor the data from the Noklus surveys ($n=294$) were normally distributed ($p < 0.001$), thus medians and non-parametric tests were used. Since the median is robust to outliers, no outliers were excluded.

Data from 12 Labquality- and six Noklus EQAMs used during a one-year period were grouped and aggregated for each MP peer group in the two schemes, respectively. The MP peer groups were Abbott Alinity, Abbott Architect, Roche Cobas, and Siemens Advia. The corresponding 95% percentile bootstrap confidence intervals (PBCIs) were calculated by sorting 10,000 median % bias bootstrap samples from smallest to largest, defining the lower and upper limits of the 95% PBCIs as the 250th and 9750th median % bias bootstrap samples, respectively (i.e., 2.5th and 97.5th percentiles) [15]. Variance homogeneity is a prerequisite for aggregation of data [16]. Therefore, the Fligner-Killeen test was employed to test for variance homogeneity in all 12 Labquality- and

all six Noklus EQAMs used within each MP, indicating that the variances were not significantly different ($p > 0.99$). Also, median % bias for each EQAM per MP was calculated. The relative biases were consistent across control levels (Supplemental Figures 1–4). All test results were unaffected by excluding the two Labquality EQAMs with the highest glucose concentrations (Table 1), indicating that these samples did not have a significant impact on the test outcome.

The non-parametric Kruskal-Wallis test was performed to detect group differences, and the Wilcoxon Rank Sum test was used for pairwise comparisons [15]. To compensate for the possible higher risk of Type 1 error in multiple hypothesis testing, Bonferroni corrected p-values were calculated by multiplying the uncorrected p-value by the number of comparisons performed [17]. With four MPs, the number of paired comparisons of % bias between MPs within each scheme was six, and the number of paired comparisons between each MP and zero, i.e., target value, was four. Also, the number of paired comparisons between the processed and minimally processed Labquality EQAMs for each MP was four. All statistical tests were two-sided, and a Bonferroni corrected p-value of 0.05 or less was considered statistically significant.

Ethical considerations

The blood donors had all provided informed consent [18]. Ethical approval was not required for this quality assurance survey.

Results

Data from nine surveys carried out from May 2021 to June 2022 were included: six from Labquality and three from Noklus. The glucose concentrations in the EQAMs used varied from 2.86 to 14.45 mmol/L (Table 1). The 58 participating laboratories' response rates varied from 66 to 98% (Table 1).

When aggregating data from all surveys in each EQA scheme, the median % biases observed for each MP in the Labquality scheme were significantly larger than in the Noklus scheme, and the difference varied from 1.2 to 4.4 percentage points ($p < 0.001$) (Figure 1). Similar median % biases were observed in a sensitivity analysis, using data only from the 25 laboratories that reported glucose results in all surveys (Supplemental Figure 5).

The median % biases observed for each MP were within target value $\pm 5\%$ in both schemes (Figure 1). In the Labquality scheme, the median % bias observed for Roche Cobas (2.9%) was significantly larger than that for Abbott Alinity (1.4%) ($p < 0.001$) and Abbott Architect (2.2%) ($p = 0.013$), and median % bias for Siemens Advia (3.2%) was significantly larger than for Abbott Alinity (1.4%) ($p = 0.003$). Also, the peer group medians for the MPs were significantly larger than the target value ($p < 0.001$) (Figure 1). In the Noklus scheme, the median % bias observed for Abbott Alinity (-0.1%), Abbott Architect (-0.5%), and Siemens Advia (-1.2%) were not significantly different from one

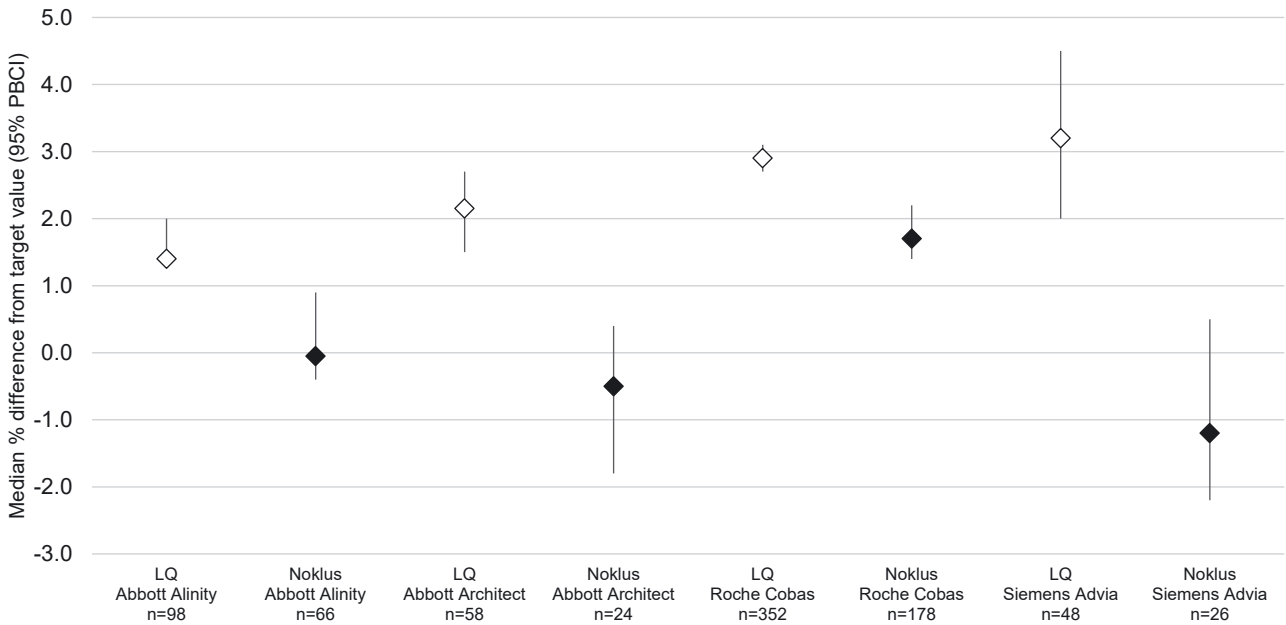


Figure 1: Aggregated EQA data from six Labquality (LQ) surveys (white) and three Noklus surveys (black). Median percent difference from target value (% bias) with 95 % percentile bootstrap confidence interval (PBCI) for four MPs. n, number of results.

other ($p > 0.270$) and not significantly different from the target value ($p > 0.756$). In contrast, the median % bias for Roche Cobas (1.7 %) was significantly larger than the other MPs ($p < 0.001$), and significantly larger than the target value ($p < 0.001$) (Figure 1). The largest difference in median biases (4.4 percent points) was observed for Siemens Advia with a bias of 3.2 % (95 % PBCI 2.0, 4.5) in the Labquality scheme and -1.2 % (95 % PBCI -2.2, 0.5) in the Noklus scheme (Figure 1). The order of bias size for the various MPs was

different in the two schemes. In the Labquality scheme, Siemens Advia had the highest bias, followed by Roche Cobas, Abbott Architect and Abbott Alinity. Whereas in the Noklus scheme, the order from highest to lowest was Roche Cobas, Abbott Alinity, Abbott Architect and Siemens Advia (Figure 1).

No significant differences in median % biases between the processed and minimally processed Labquality EQAMs were found ($p > 0.094$) (Figure 2).

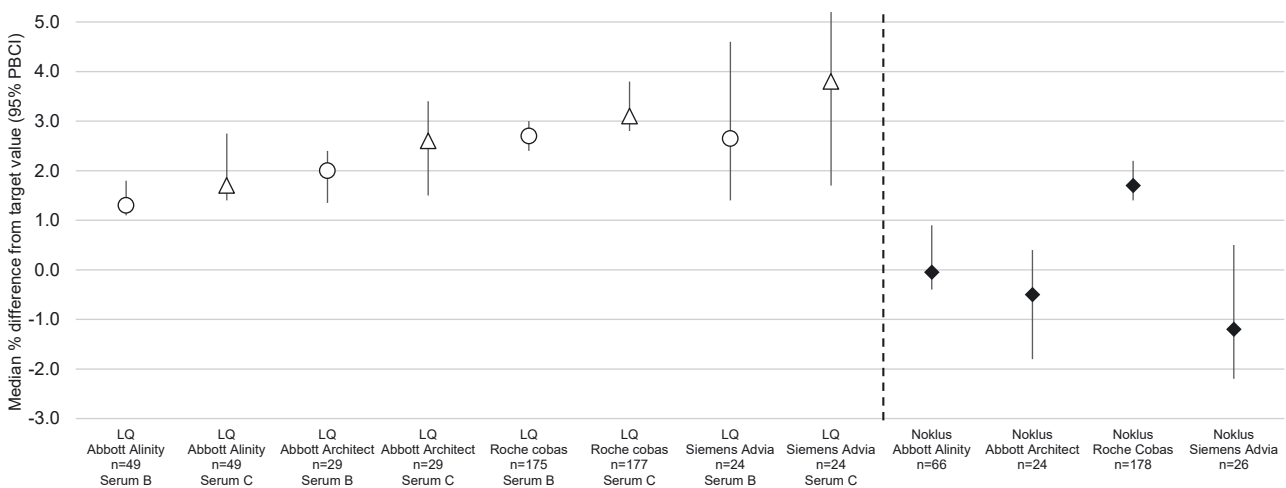


Figure 2: Aggregated EQA data from EQA materials serum B (minimally processed) and serum C (processed) measured in six Labquality (LQ) surveys (white), and EQA materials produced by Noklus measured in three Noklus surveys (black). Median percent deviation from target value (% bias) with 95% percentile bootstrap confidence interval (PBCI) for four MPs. n, number of results.

Discussion

In this study, aggregated data from two providers of EQA schemes for serum glucose showed different biases between MPs and compared to the target value. In the Labquality scheme, all four MPs had positive biases that were larger than in the Noklus scheme. Roche Cobas had a larger bias than Abbott Alinity and Abbott Architect, and Siemens Advia had a larger bias than Abbott Alinity. In the Noklus scheme, glucose measurements were equivalent between Abbott Alinity, Abbott Architect and Siemens Advia, while Roche Cobas had a larger bias. The median percent biases with 95 % PBCIs observed for each MP were within acceptance limits in both schemes.

The two prerequisites for demonstrating the metrological traceability of an MP through EQA are the use of (1) verified commutable EQAMs, and (2) commutable certified reference material or a reference MP [3]. The EQAMs used in the Labquality scheme have not been formally assessed for commutability, neither the minimally processed nor the processed materials. No significant differences in median biases between minimally processed and processed EQAMs were found (Figure 2). This indicates that the minimally processed and processed Labquality EQAMs are equally commutable- or non-commutable. The RSX that is used in the target value determination has not been assessed for commutability for glucose, probably since both RSX and its predecessor IMEP-17 were minimally processed, and the production processes were designed to ensure commutability [9]. The Noklus glucose EQA scheme on the other hand, uses commutable EQAMs and target values that are determined by the IDMS highest-order reference MP. This approach avoids any transfer bias in the value determination from higher-order reference materials to the EQAMs used by the EQA scheme. Thus, the prerequisites for demonstrating the metrological traceability of an MP through EQA are met [3], and is probably why the biases are smaller in the Noklus scheme than in the Labquality scheme. The difference between results obtained in the two EQA schemes could be due to non-commutability of EQAMs circulated by Labquality, or non-commutability of the RSX [3].

Of the MPs included in this study, all except Abbott Architect were included in the Noklus commutability study. However, there is a high probability that the commutability results are also valid for this MP since the Abbott Alinity system was developed to provide comparable results to the Architect systems, demonstrated for glucose in a previous study [19]. This assumption is in line with the recommendations from the IFCC working group on commutability [6].

The optimal is to include as many different MPs and analytical measurement principles as possible in the commutability assessment. However, the working group acknowledges that it may not be possible to include all MPs, and in such cases, they propose including the most representative group of MPs to improve the likelihood of a material being suitable for use with other MPs not included in the commutability assessment. This approach makes it more feasible for larger schemes with multiple MPs to conduct commutability studies. Furthermore, the working group has recently submitted a paper [20] that in detail describes the criterion for commutability and how commutability studies can be performed for EQAMs. An application has been developed from this work, which makes it easier to elaborate on commutability studies for many MPs and EQAMs.

As can be seen from Figure 1, the order of bias size for the various MPs was different in the two schemes, indicating that it is not only the target value that differs. The largest difference in biases between the two schemes was observed in the Siemens Advia group, with a median bias of 3.2 % in the Labquality scheme and -1.2 % in the Noklus scheme, corresponding to a difference of 0.24 and 0.44 mmol/L at glucose concentrations 5.5 and 10 mmol/L, respectively. The median bias of 3.2 % observed for the Siemens Advia group in the Labquality scheme was significantly larger than the target value, while in the Noklus scheme the observed bias of -1.2 % was not significantly different from the target value, meaning that the Siemens Advia has no bias. Consequently, if one laboratory using Siemens Advia receives feedback indicating that it produces unbiased results in the Labquality scheme (i.e., 3.2 % lower than the median), it may indeed measure glucose concentrations that are actually 3.2 % lower than the true value. This could, at least on a population level, lead to a delay in diagnosis of diabetes.

It has been argued that EQA organizations should share results from schemes using commutable EQAMs to provide large number of results that can be used in post-market surveillance [21]. However, lack of documentation of commutability of the EQAMs is a major challenge [22]. Thienpont et al. found that even minimal processing of serum could compromise its commutability [23]. Miller et al. argue that laboratories should, if possible, participate at least annually in schemes supplying commutable EQAMs, and that EQA providers should develop schemes with commutable EQAMs whenever possible [24]. Currently, Norwegian hospital laboratories participate twice a year in this type of EQA scheme for glucose.

A limitation of our study is that some MPs have few results. Nonetheless, the aggregation of EQA data improves reliability of the data summary and increases the likelihood that it meets the assumptions necessary for statistical analyses.

Strengths of this study include that the same laboratories were participating in both schemes, and thereby the MPs could be grouped identically. Previous studies have described challenges regarding the large variation in how the different EQA providers register manufacturer details [22, 25]. Another strength was that the data was collected from both schemes during the same time period.

We have shown that laboratories participating in EQA schemes from two EQA providers, with different EQA designs, obtain different glucose results. This underlines the importance of EQA providers using verified commutable EQAMs and target values traceable to reference measurement procedures in EQA schemes designed for assessment of the metrological traceability of laboratory results.

Research ethics: The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This quality assessment study was exempt from review.

Informed consent: Informed consent was obtained from all individuals included in this study, or their legal guardians or wards.

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

Competing interests: The authors state no conflict of interest.

Research funding: None declared.

Data availability: The raw data can be obtained on request from the corresponding author.

References

- American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care* 2013;36:S11–66.
- World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. In: Report of a WHO/IDF consultation 2006 [Online]. http://apps.who.int/iris/bitstream/handle/10665/43588/9241594934_eng.pdf?sequence=1 [Accessed 19 May 2023].
- Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* 2011;57:1670–80.
- Miller WG, Sandberg S. Quality control of the analytical examination process. In: Rifai N, editor. *Tietz textbook of laboratory medicine*, 7th ed. St Louis: Elsevier; 2022:129–63 pp.
- Clinical Laboratory Standards Institute (CLSI). Evaluation of commutability of processed samples; approved guideline. CLSI document EP14-A3. Wayne (PA): CLSI; 2014.
- Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC working group recommendations for assessing commutability Part 1: general experimental design. *Clin Chem* 2018;64:447–54.
- Henriksen GM, Pedersen MM, Norgaard I, Blom M, Blou L, Blaabjerg O, et al. Minimally processed fresh frozen human reference sera: preparation, testing, and application to international external quality assurance. *Scand J Clin Lab Invest* 2004;64:293–308.
- International Organization for Standardization. *Statistical methods for use in proficiency testing by interlaboratory comparison*, ISO 13528. Geneva, Switzerland: International Organization for Standardization; 2015.
- Pedersen MM, Rustad P, Simonsson P. Certificate of analysis NFKK reference serum X: a reprint. *Scand J Clin Lab Invest* 2004;64:321–6.
- Örnemark U, Uldall A, Van Nevel L, Aregbe Y, Taylor P. IMEP-17 trace and minor constituents in human serum. In: Certification report, internal report GE/R/IM/36/01 (EUR 20243 EN). Geel: IRMM; 2002.
- Dean RB, Dixon WJ. Simplified statistics for small numbers of observations. *Anal Chem* 1951;23:636–8.
- Gislefoss RE, Lauritzen M, Langseth H, Morkrid L. Effect of multiple freeze-thaw cycles on selected biochemical serum components. *Clin Chem Lab Med* 2017;55:967–73.
- Cuhadar S, Koseoglu M, Atay A, Dirican A. The effect of storage time and freeze-thaw cycles on the stability of serum samples. *Biochem Med* 2013;23:70–7.
- The Joint Committee for Traceability in Laboratory Medicine (JCTLM). Traceability, education and promotion [Online]. <https://jctlm.org/> [Accessed 19 May 2023].
- Devore J, Berk K, Carlton M. *Modern mathematical statistics with applications*. In: Springer texts in statistics, 3rd ed. Cham, Switzerland: Springer; 2021.
- Conover WJ, Johnson ME, Johnson MM. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 1981;23:351–61.
- Lydersen S. Adjustment of p-values for multiple hypotheses. *Tidsskr Nor Laegeforen* 2021;141:1292.
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.
- Krintus M, Fernandez JA, Chesters C, Colla R, Ford C, Frattolillo D, et al. Analytical performance of 10 high-volume clinical chemistry assays on the Alinity c system. *Lab Med* 2019;50:e1–8.
- Sandberg S, Fauskanger P, Johansen JV, 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. submitted.
- Badrick T, Stavelin A. Harmonising EQA schemes the next Frontier: challenging the status quo. *Clin Chem Lab Med* 2020;58:1795–7.
- van der Hagen EAE, Weykamp C, Sandberg S, Stavelin AV, MacKenzie F, Miller WG. Feasibility for aggregation of commutable external quality assessment results to evaluate metrological traceability and agreement among results. *Clin Chem Lab Med* 2020;59:117–25.
- Thienpont LM, Van Uytendaele K, Marriott J, Stokes P, Siekmann L, Kessler A, et al. Feasibility study of the use of frozen human sera in split-sample comparison of immunoassays with candidate reference measurement procedures for total thyroxine and total triiodothyronine measurements. *Clin Chem* 2005;51:2303–11.
- Miller WG, Myers GL, Lou Gantzer M, Kahn SE, Schonbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 2011;57:1108–17.
- 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.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/cclm-2023-0532>).