



HAL
open science

Multilevel qualification of a large set of blood gas analyzers: Which performance goals?

Nils Kuster, Anne-Sophie Bargnoux, Stéphanie Badiou, Anne-Marie Dupuy,
Laurence Piéroni, Jean-Paul Cristol

► To cite this version:

Nils Kuster, Anne-Sophie Bargnoux, Stéphanie Badiou, Anne-Marie Dupuy, Laurence Piéroni, et al.. Multilevel qualification of a large set of blood gas analyzers: Which performance goals?. *Clinical Biochemistry*, 2019, 10.1016/j.clinbiochem.2019.09.005 . hal-02291754

HAL Id: hal-02291754

<https://hal.science/hal-02291754>

Submitted on 7 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multilevel qualification of a large set of blood gas analyzers: Which performance goals?

Nils Kuster^{a,b}, Anne Sophie Bargnoux^{a,b}, Stéphanie Badiou^{a,b}, Anne-Marie Dupuy^a, Laurence Piéroni^{a,1}, Jean Paul Cristol^{a,b,*}

^a Department of Biochemistry and Hormonology, CHU de Montpellier, France

^b PhyMedExp, University of Montpellier, INSERM, CNRS, France

A B S T R A C T

Keywords:

Blood gas
Analytical performance
Biological variations

Background: Blood gas analyzers are frequently installed as point of care devices and thus allow rapid decision making. Few data are available regarding analytical performance of large sets of BGA. We aimed at evaluating 22 ABL 90 Flex Plus analyzers intended to be deployed. The evaluation was performed at the device level and at the entire set level to characterize the quality of measurements but also to ensure consistency across the devices deployed in the hospital.

Methods: Imprecision and total error were assessed for pH, pCO₂, pO₂, sodium, potassium, ionized calcium, glucose, lactate and oximetry parameters. Imprecision at the hospital level including between device variability was also evaluated. One of the two analyzers used in the central laboratory was correlated with a GEM Premier 4000 and a Cobas b221 analyzers. Thereafter, we tested sequentially the 20 instruments intended to be deployed in care service in comparison with the reference device.

Results: Heterogeneity of analytical performance across the different analyzers was low, allowing to consider the whole set as a unique analyzer. The total error was in line with performance goals. Analytical performance of the analyzers was found suitable for use in clinical practice.

Conclusions: Our study is an example of the qualification of a set of point and underscores 1) The need for a unified qualification scheme when multiple analyzers are deployed simultaneously 2) analytical performance goals compatible with clinical use and the state of the art for all parameters.

1. Introduction

Blood Gas (BG) analysis can be of major importance in critically ill patients [1]. Point of Care (POC) devices allow extra laboratory, near-patient analysis and rapid decision making. In the last few years, many parameters beyond blood gas analysis (pH, pCO₂, pO₂) have been implemented and modern analyzers yield a measure of electrolytes (Na⁺, K⁺), ionized calcium (Ca²⁺), metabolites (glucose, lactate) or oximetry parameters (total hemoglobin (tHb), oxyhemoglobin(O₂Hb), carboxyhemoglobin(COHb), methemoglobin, (MetHb)) [2]. Quality assurance is becoming an essential part of laboratory medicine, guided by international standards, such as ISO 15189:2012 – Medical laboratories – Requirement for quality and competence. In the field of point of care testing, ISO 22870:2016 completes these requirements. In order to

meet the need for quality assurance, analytical performance of analyzers must be verified before they can be used in clinical practice. Some data are available regarding the analytical performance of many types of BG analyzers (BGA), including the ABL 90 Flex Plus (Radiometer®, Copenhagen, Denmark) [3–5]. However, few data are available regarding analytical variability between different analyzers of the same model. In large hospitals, BGA are frequently installed as a set of analyzers. It is thus interesting to ensure consistency of results across all devices of a hospital. The aim of this study is to evaluate within and between-device variability in a set of 22 ABL 90 Flex Plus analyzers.

2. Material and methods

This study aimed at qualifying blood gas parameters (pH, pCO₂,

Abbreviations: BG, Blood Gas; POC, Point of Care; tHb, total hemoglobin; TE, Total Error; LOA, Limit of Agreement; BV, Biological Variation; O₂Hb, oxyhemoglobin; COHB, Carboxyhemoglobin; MetHb, Methemoglobin

* Corresponding author at: Department of Biochemistry, Montpellier University Hospital, 371, avenue du doyen Gaston Giraud, 34090 Montpellier, France.

E-mail address: jp-cristol@chu-montpellier.fr (J.P. Cristol).

¹ These authors contributed equally to this work.

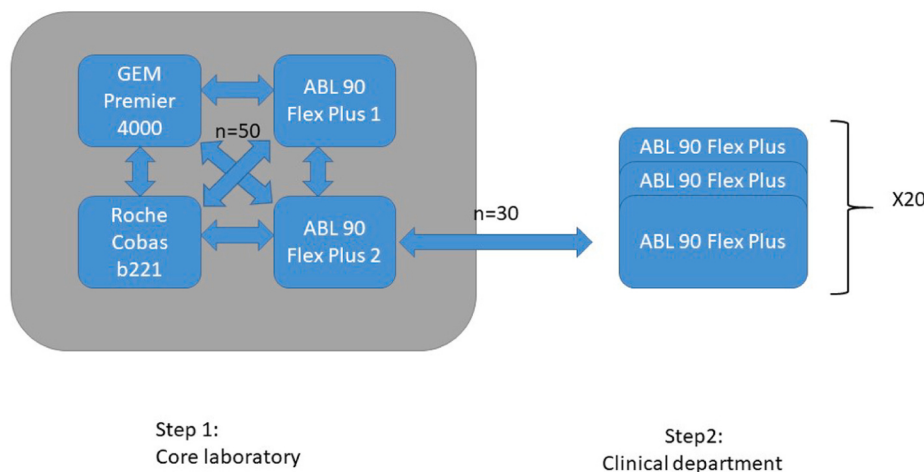


Fig. 1. Qualification scheme used for correlations analysis.

pO₂), electrolytes (Na⁺, K⁺), ionized calcium (Ca⁺⁺), metabolites (glucose and lactate) and oximetry parameters (tHb, O₂Hb, COHb, MetHb). The ABL 90 Flex Plus is a recent, easy-to-use, cartridge-type analyzer [5]. Using this analyzer, pH, pCO₂, Na⁺, K⁺, Ca⁺ are measured using potentiometry methods, glucose and lactate using amperometry method. tHb, O₂Hb, COHb, MetHb are measured by spectrophotometry. Lastly, pO₂ is measured by optical sensing using phosphorescence quenching [6].

Between October 2017 and February 2018, 20 ABL 90 Flex Plus analyzers have been tested and deployed in 10 departments of the University Hospital of Montpellier and the Institut du Cancer de Montpellier (ICM). Two additional devices were set up in the central laboratory to replace three Cobas b221 (Roche diagnostics) previously used.

2.1. Metrological traceability

The Metrology Section at Radiometer is responsible for establishing metrological traceability for the parameters pH, pCO₂, pO₂, Na⁺, K⁺, Ca⁺⁺, glucose, lactate, tHb, O₂Hb, COHb, MetHb [7].

pH is traceable to Standard Reference Materials (SRM) of the National Institute of Standards and Technology (NIST) g186I/II-g, 185 g, 187e, 191-I-d, 191-II-d. Secondary reference method is the ABL 735 Analyzer. Traceability of pCO₂ and pO₂ is ensured by tonometry using NIST SRM 1674b and 2658a. Traceability of Na⁺ and K⁺ is performed according to NIST SRM 919b (NaCl) and 999b (KCl). ABL 735 analyzer is used as a secondary reference method. Traceability of Ca⁺⁺ is performed using reference material NIST SRM 915b, ABL 735 is used as a secondary reference method. Traceability of glucose is performed using reference material NIST SRM 917c (D-glucose) and spectrophotometry. Traceability of lactate uses reference solutions prepared from pure lactic acid provided by Sigma Chemical Company. Primary reference material of tHb is prepared using a spectrophotometric method based on NIST SRM 930D. COHb traceability is based on a 100% COHb blood sample prepared using certified 99.997% CO. MetHb traceability is obtained through a primary blood sample prepared by addition of KCN according to Evelyn and Malloy. Primary blood sample are used for calibration of ABL 735 reference analyzer for all oximetry parameters.

2.2. Qualification scheme

This study has been approved by the Institutional Review Board of the Montpellier University Hospital (number 2018_IRB_MTP_12-13).

Analytical performance studies have been performed by trained lab staff regarding ABL 90 Flex Plus used in a POC setting or as an in-lab

blood gas analyzer.

According to ISO 15189 norm, all devices were tested for bias, imprecision, and correlation as follows:

2.2.1. Total error in repeatability conditions

The within-run imprecision (repeatability) of all devices was tested through the measurement of the external quality controls (Qualicheck 5 +) at three levels. The three levels have been measured 30 times each using control material vials provided by Radiometer (total: 30 × 3 vials/apparatus measured the same day).

Since Qualicheck solutions have assigned values for each parameter at each level, we consider the difference between the means of measurements and assigned levels as a surrogate of the bias. We thus computed bias, imprecision and total error (TE) in these repeatability conditions.

At the set of devices level, mean of all measurements and CV were calculated, allowing estimation of total error at each of the three levels.

2.2.2. Reproducibility

Between-run imprecision (reproducibility) was assessed using internal control solutions. Three control solutions are embedded in the ABL 90 Flex Plus solutions pack. Internal controls were scheduled twice a day, for 15 days (30 measurements of each level). Twelve reagent lots were used during the qualification period, (1 to 4 lots per apparatus).

2.2.3. Correlations

One ABL 90 Flex Plus intended for use in core laboratory device was set as a reference device and was correlated against previous apparatuses used in core laboratory (Cobas b221) and in clinical departments (GEM Premier 4000). Since one unit is only used for demonstration purpose during the training of medical practitioners, this unit has not been used in the correlation analysis, leading to 21 apparatuses considered in the correlation analysis.

Following these initial correlations, we used a transitive model in which apparatuses intended for use in clinical departments were correlated with the reference apparatus in the core laboratory. Correlations were performed using 50 samples for the devices in the core laboratory and 30 samples for the devices in clinical departments (Fig. 1). Glucose and lactate parameters were not available on Cobas b221 and GEM Premier 4000 analyzers and were thus correlated against a Cobas 8000 (Roche, Meylan, France) apparatus.

2.3. Statistical analysis

2.3.1. Repeatability, total error and measurement uncertainty

In repeatability analysis, standard deviation (*sd*) and mean (*m*) were

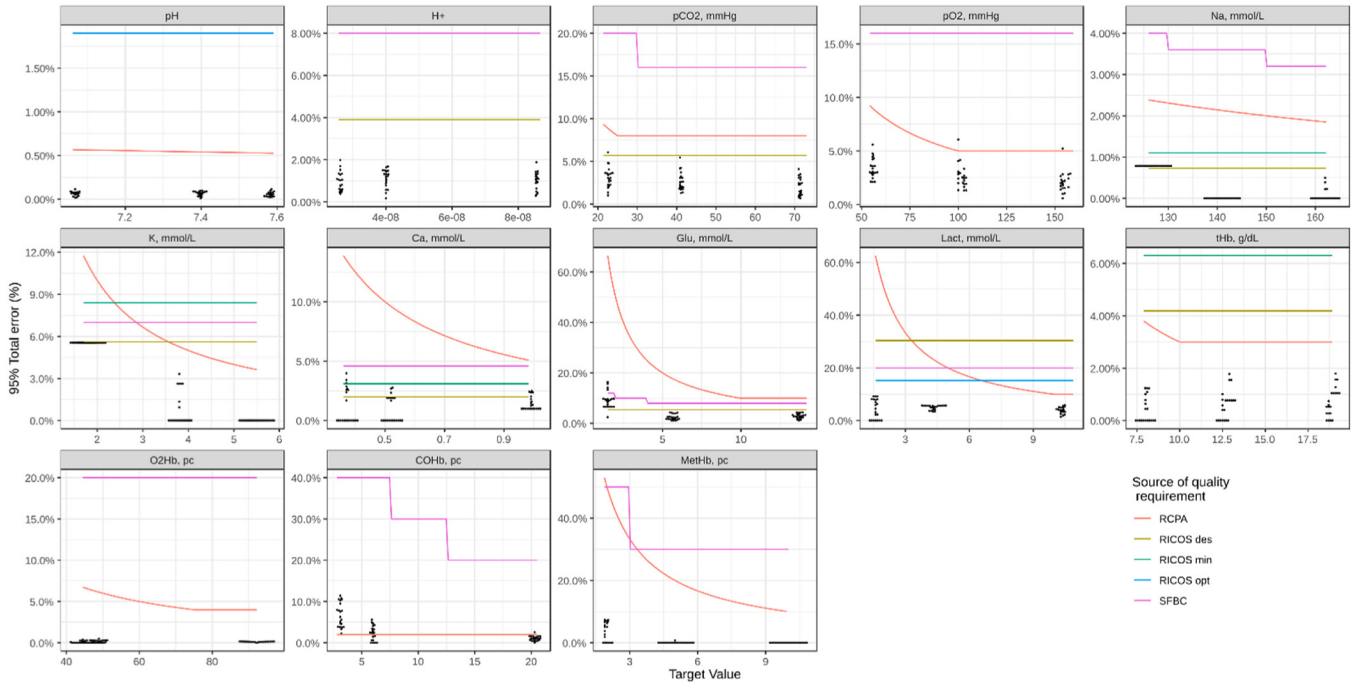


Fig. 2. Total error (%) at the device level at each level of each parameter considered in the evaluation process. Total error goals are depicted as coloured lines according to the specifications database, RCPA: Royal College of Pathologists of Australasia, RICOS des: RICOS desirable specifications from biological variations database, RICOS min: RICOS minimal specifications from biological variations database, RICOS opt: RICOS optimal specifications from biological variations database, SFBC: specifications from the Société Française de Biologie Clinique.

derived from the 30 measurements at each level, parameter, and device. Coefficients of variation of the parameters were computed as follows:

$$CV(\%) = \frac{sd}{m}$$

Bias was defined as the mean of the 30 measurements minus the target value assigned by the manufacturer

$$(\text{Bias} =) m - \text{target}$$

Expressed as a percentage, bias was defined as:

$$\text{Bias}(\%) = \frac{\text{bias}}{\text{target}} \times 100$$

The total error at the 95% level was computed using a simulation method. It was estimated according to bias and imprecision as the 95th percentile of the absolute value of a normally distributed sample of size $n = 10,000$ with mean = bias and standard deviation = sd. Total error was also expressed as a percentage of the target value and compared to analytical performance goals. In addition, measurement uncertainty was also computed using proficiency testing materials (The method is described in Supplementary data 1)

2.3.2. Reproducibility

Reproducibility was assessed using internal control solutions. Assigned target levels depend on cartridge lot. Thus, measurements results were centred prior to analysis according to cartridge lot number to avoid introducing a systematic difference in the evaluation of imprecision. For each analyzer and parameter, standard deviation of measurements was computed as follows:

$$sd^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_{ij} - m_j)^2$$

where x_{ij} represents the i^{th} measurement for lot j , m_j is the mean of measurements for lot j and n is the total number of measurements.

Analysis of variance (ANOVA) was used to evaluate within and between-device components of the total variance.

2.3.3. Correlations

Correlation data were analyzed through Passing-Bablok regression. Agreement between the two methods was assessed according to the method described by Bland and Altman [8]. In order to compute the limits of agreement for the whole set of analyzers, we used mixed effects limits of agreement based on the method proposed by Parker et al. [9]. This approach allows taking into account the clustered structure of the data. Mixed-effects models were computed for each parameter, considering analyzers as a random effect. We thus considered the set of analyzers as a sample of size 21 from a larger population of analyzers. In this context, the paired difference between two measurements from two devices, Y_{ij} is modelled as follows:

$$Y_{ij} = \alpha + r_j + e_{ij}$$

with $r_j \sim N$) et $e_{ij} \sim N$) where α is the constant intercept term, r_j is the random effect of the j^{th} analyzer and e_{ij} the error term of the i^{th} measurement using the j^{th} analyzer

According to CLSI EP09-A3 protocol, outliers were identified via visual inspection of the data and the Extreme Studentized deviate (ESD) technique [10]. Bland and Altman bias and limits of agreement were also computed without discarding outliers.

2.4. Analytical performance goals

Multiple sources of analytical performance goals were considered for comparison purpose. Ricos' database [11,12] was used preferably. Some parameters were not available in Ricos' database (pO2 and oximetry in particular) or were evaluated in an old study (pH). RCPA (Royal College of Pathologists of Australasia) database [13] for total error and SFBC (French society of clinical biology) database [14] for imprecision goals were considered in these cases.

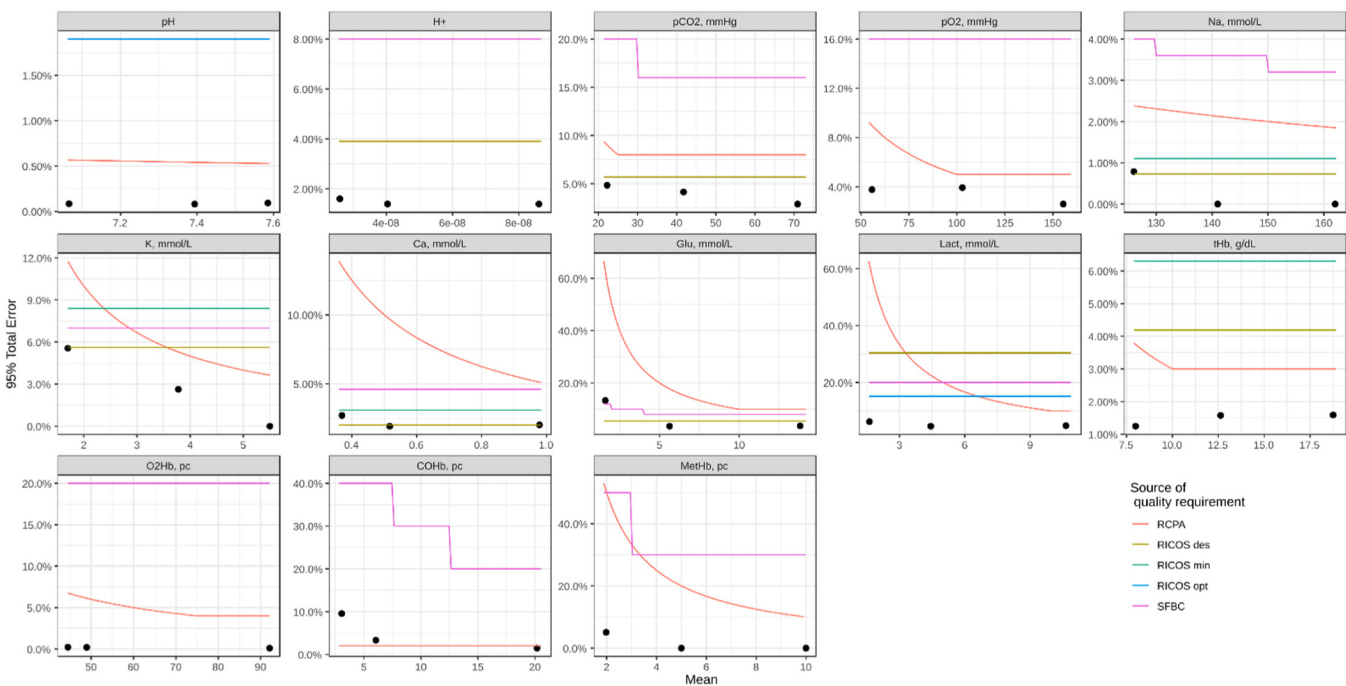


Fig. 3. Total error (%) at the whole system level at each level of each parameter considered in the evaluation process. Total error goals are depicted as coloured lines according to the specifications database, RCPA: Royal College of Pathologists of Australasia, RICOS des: RICOS desirable specifications from biological variations database, RICOS min: RICOS minimal specifications from biological variations database, RICOS opt: RICOS optimal specifications from biological variations database, SFBC: specifications from the Société Française de Biologie Clinique.

3. Results

3.1. Total error in repeatability conditions

Total errors for each level of each parameter and each device are shown in Fig. 2. Allowable total error according to parameter value and database of analytical performance requirements are depicted as coloured lines.

The performance characteristics of all devices were compliant with the RCPA quality requirements at all levels except for pO₂ for one apparatus, and COHb for multiple analyzers. Considering Ricos et al. database requirements, some devices did not comply with minimal specifications at the lowest level of ionized calcium ($n = 3$). Ricos et al. database does not include quality requirements for pO₂ and oximetry parameters. RCPA database allows assessment of performance for all parameters of blood gas, electrolytes and oximetry parameters and leads to overall performance compatible with the needs in a clinical environment, with the exception of multiple devices for COHb.

At the entire set of apparatuses level (Fig. 3), the set complied with RCPA recommendations at all parameters and levels except for COHb (for 2 levels). The set also complied with SFBC requirements except for the lowest level of glucose (1.5 mmol/L).

3.2. Reproducibility

Imprecision Cvs are presented in Table 1. The total CV was computed across all apparatuses and broken down into within and between device imprecision. Ricos et al. desirable performance database was used as imprecision goals since it provides imprecision goals for each parameter except pH, pO₂, O₂Hb, COHb and MetHb, for which SFBC recommendations were considered.

Imprecision at the whole set level was in line with analytical performance goals with the exception of ionized calcium at the lowest level (mean of 0.31 mmol/L) and the lowest level of pO₂ (mean = 17.5 mmHg). The set of analyzers can thus be evaluated as one unique system providing results in the different clinical departments.

3.3. Correlation and agreement analysis

Correlation and agreement of the reference ABL90 Flex Plus deployed in the central laboratory and Cobas b221 and GEM Premier 4000 are presented in Table 2. O₂Hb, COHb and MetHb were not available on Cobas b221 and GEM Premier 4000 devices and thus could not be correlated against previous apparatuses.

Passing and Bablok correlations of ABL 90 Flex Plus against Cobas b221 led to slopes between 0.933 (Na) and 1.222 (tHb). These slopes were compensated by the intercepts (8.167 and - 1.644, respectively). Bias, as evaluated according to Bland and Altman's method, was acceptable for all parameters

Results of agreement analysis according to Bland and Altman's method for the 20 units against one device in the central laboratory is shown in Fig. 4. These results show a strong homogeneity in results across the twenty devices tested. For pH analysis median bias was 0 and limits of agreement were very close to 0. Limits of agreement (LOA) across all devices (at the apparatus level) were computed and represented at the set level in Fig. 3. For pCO₂, the lowest lower LOA was - 3.29 and highest upper LOA was 1.19 mmHg, indicating a very strong agreement between the different devices. For pO₂ analysis, median bias was - 0.44 mmHg. One device presented particularly large limits of ranging from - 4.06 to + 5.78 mmHg. For Na⁺ measurement, only one device had limits of agreement exceeding ± 2 mmol/L. For K⁺ measurement, all devices but one had limits of agreement in the range ± 0.2 mmol/L. For ionized calcium, lowest lower LOA was - 0.02 mmol/L and highest upper LOA was 0.03 mmol/L. For lactate measurement, lowest LOA was - 0.14 mmol/L and highest upper LOA was 0.43 mmol/L. The most important heterogeneity of results was observed for tHb. Two apparatuses showed limits of agreement wider than ± 20 g/L. Further investigation revealed that total hemoglobin measurement outliers were probably linked to an insufficient homogeneity of the sample.

Table 1

Within device, between device and total imprecision of the 22 devices. Analytical performance goals were derived from RICOS at all desirable performance goals, except for pH, pO₂, O₂Hb%, COHb%, and MetHb %. For these parameters, SFBC goals were considered.

Level	Parameter	Mean	Within device CV	between device CV	total CV	Goal	Specifications database
pH [H +],unit (nmol/L)	Low	6.77(169.2)	0.635%	0.537%	0.832%	1.5%	SFBC
	Medium	7.15(70.5)	0.512%	0.229%	0.561%	1.5%	SFBC
	High	7.49(32.5)	0.582%	0.655%	0.876%	1.5%	SFBC
pCO ₂ , mmHg	Low	20.19	1.189%	0.908%	1.496%	2.4%	RICOS (desirable)
	Medium	28.35	0.882%	0.610%	1.072%	2.4%	RICOS (desirable)
	High	65.98	0.756%	0.880%	1.160%	2.4%	RICOS (desirable)
pO ₂ , mmHg	Low	17.49	4.181%	3.149%	5.234%	1.5%	SFBC
	Medium	188.92	0.560%	0.288%	0.630%	1.5%	SFBC
	High	302.06	0.299%	0.147%	0.334%	1.5%	SFBC
Na, mmol/L	Low	112.20	0.207%	0.064%	0.217%	^a 0.3%	RICOS (desirable)
	Medium	135.80	0.251%	0.072%	0.261%	^a 0.3%	RICOS (desirable)
	High	171.10	0.172%	0.047%	0.178%	^a 0.3%	RICOS (desirable)
K, mmol/L	Low	1.74	1.212%	0.792%	1.448%	^a 2.3%	RICOS (desirable)
	Medium	3.73	0.526%	0.033%	0.527%	^a 2.3%	RICOS (desirable)
	High	6.51	0.168%	0.032%	0.171%	^a 2.3%	RICOS (desirable)
Ca, mmol/L	Low	0.31	1.238%	0.930%	1.549%	^a 0.9%	RICOS (desirable)
	Medium	0.81	0.469%	0.000%	0.469%	^a 0.9%	RICOS (desirable)
	High	1.54	0.294%	0.110%	0.313%	^a 0.9%	RICOS (desirable)
Glu, mmol/L	Medium	6.78	0.831%	1.324%	1.563%	^a 2.8%	RICOS (desirable)
	High	14.58	0.637%	0.699%	0.945%	^a 2.8%	RICOS (desirable)
Lactate, mmol/L	Medium	3.50	1.123%	0.804%	1.381%	13.6%	RICOS (desirable)
	High	7.49	0.967%	1.302%	1.622%	13.6%	RICOS (desirable)
tHb, g/L	Medium	129.2	0.239%	0.274%	0.363%	1.43%	RICOS (desirable)
	High	192.8	0.120%	0.193%	0.227%	1.43%	RICOS (desirable)
O ₂ Hb, %	Medium	48.96	0.093%	0.069%	0.116%	3.8%	SFBC
	High	92.22	0.056%	0.021%	0.060%	3.8%	SFBC
COHb, %	Medium	3.69	2.007%	3.149%	3.734%	7.5%	SFBC
	High	20.09	0.261%	0.453%	0.523%	3.8%	SFBC
MetHb, %	Medium	1.78	2.180%	1.953%	2.927%	11.3%	SFBC
	High	10.00	0.000%	0.000%	0.000%	7.5%	SFBC

Imprecision exceeding goals are in bold.

^a goals for serum parameters were used.

Table 2

Correlation and Bland and Altman analysis between the reference ABL 90 Flex Plus in the central laboratory and previously used Cobas b221 (A) and GEM Premier 4000 (B).

Parameter	Passing Bablok regression		Bland and Altman analysis		
	Slope	Intercept	Mean difference	Lower limit of agreement	Upper limit of agreement
A	Cobas b221				
pH	0.917	0.599	-0.018	-0.04	0.004
pCO ₂ (mmHg)	1.077	-2.7	0.6	-2.2	3.3
pO ₂ (mmHg)	1.004	3.2	3.5	-4.2	11.1
tHb (g/L)	1.222	-16.44	7.74	-8.67	24.15
Na (mmol/L)	0.933	8.2	-1.1	-2.7	0.6
K (mmol/L)	1.018	-0.204	-0.2	-0.4	0.1
Ca ²⁺ (mmol/L)	1.022	-0.03	0	-0.04	0.03
B	GEM Premier 4000				
pH	1.071	-0.531	-0.001	-0.024	0.021
pCO ₂ (mmHg)	1.059	-2.292	0.4	-2.3	3.1
pO ₂ (mmHg)	0.979	-1.013	-2.9	-9.5	3.7
tHb (g/L)	1	3.00	3.40	-7.41	14.22
Na (mmol/L)	1	-3	-2.9	-4.6	-1.1
K (mmol/L)	1	0	0.0	-0.3	0.3
Ca ²⁺ (mmol/L)	1.05	-0.05	0.01	-0.02	0.04

4. Discussion

Only a few data are available regarding variability in blood gas results across all devices deployed in a hospital. The replacement of a set of 22 analyzers offered an occasion to evaluate the variability of analytical performance across a set of more than 20 ABL 90 Flex Plus. We could also correlate ABL 90 Flex Plus to Cobas b221 and GEM

Premier 4000 analyzers.

4.1. Homogeneity of analytical performance

Although recent studies have reported data regarding the installation of a set of analyzers [4,15], most studies dedicated to assessing analytical performance of blood gas analyzers use only one analyzer of each type [16]. These studies are very important, particularly for new instruments and because modern analyzers now allow measurements of a growing number of parameters at the same time. Our study clearly indicates that the performance of the set of analyzers is essentially homogeneous. Between device imprecision was found to be low and the whole set can thus be regarded as one unique system measuring samples in different places of the hospital (Table 1). However, a certain degree of heterogeneity is inevitable when considering a large number of instruments. This study clearly illustrates that between-device variation should be taken into account. It can be speculated that the cause of between-device variation lies in the production process of the devices but also in the lot-to-lot variation of reagents. As a result, some devices did not fulfil international specifications such as RICOS's or RCPC's for some levels/analytes. However, the computation of measurement uncertainty from proficiency testing materials indicates that these instruments are appropriate for use in clinical practice (Supplementary data Fig. S1). Largest imprecision has been observed for metabolites (glucose, lactate) and the lowest level of ionized calcium. The challenge in measuring low levels of ionized calcium using blood gas analyzers has already been raised in previous studies, particularly in the context of regional citrate anticoagulation [17]. In our study, total error at a low level of ionized calcium was acceptable according to minimum RICOS acceptable error derived from biological variations, however total imprecision across all devices did not fulfil minimum imprecision criterion (total CV = 1.54%). This observation confirms that

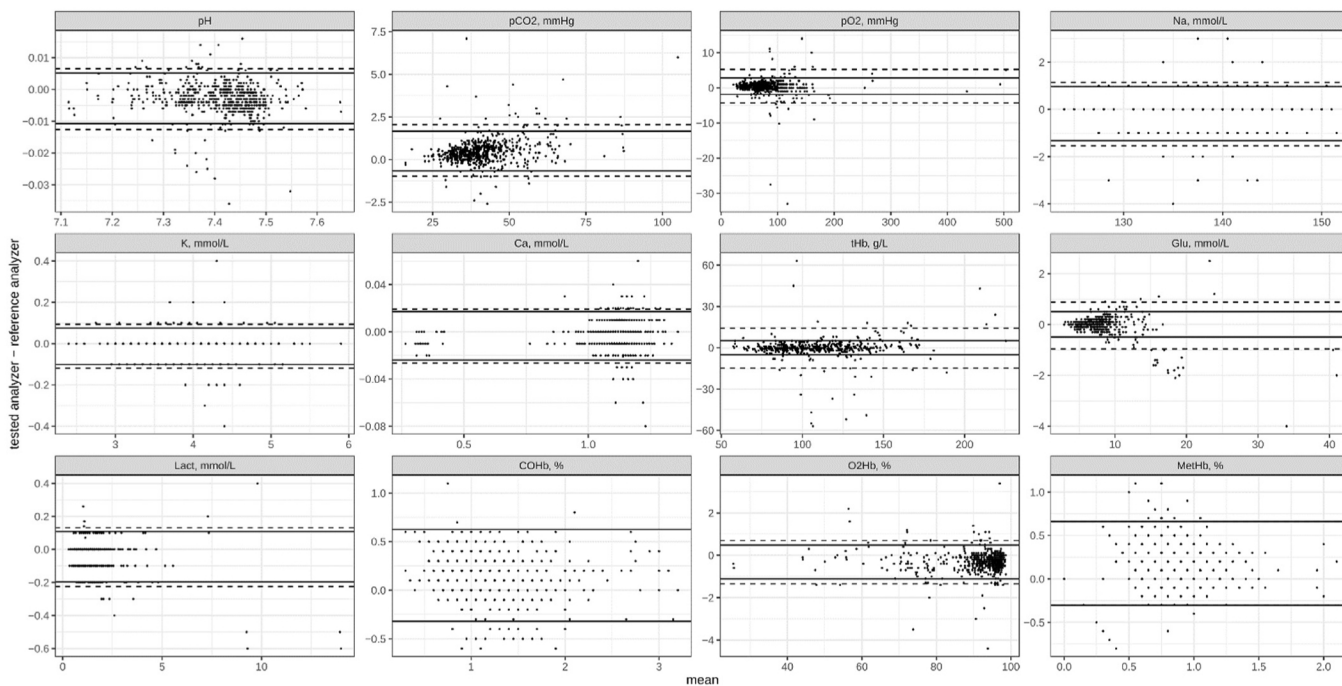


Fig. 4. Bland and Altman analysis for each parameter across all devices against the reference analyzer in central laboratory. Plain lines represent limits of agreement computed after exclusion of outliers and dashed lines represent limits of agreement computed without removing outliers.

manufacturers should maintain their effort in improving analytical performance for measuring low levels of ionized calcium, as these results are routinely used to adjust post-filter citrate administration.

4.2. Sets of analyzers require multilevel qualifications

Because hospitals tend to get bigger, it is likely that several units of POC analyzers would be deployed in different care service. The whole set of analyzers can also be seen as a unit that has to provide physicians with the same results whatever the care service. The qualification of the analyzers should be performed at the analyzer level but also at the population of analyzers level to ensure consistency at the hospital level. In our study, we computed imprecision and total error at the device level and at the whole set level.

Although some studies evaluated a set of analyzers [4,15], such verification is sparsely covered in the literature and clearly, need further studies. In our study, the set of analyzers as a whole demonstrated acceptable analytical performance (Fig. 2).

To our knowledge, no multilevel qualification scheme has been validated, which underlines the need for further studies. Large sets of analyzers generate large datasets. Specific data analysis approaches, like analysis of variance (ANOVA) or mixed-effect models, allow to treat large dataset and distinguish within-device and between-devices characteristics.

Considerable work has been done in order to provide laboratories with reliable analytical performance goals, collecting large amounts of data, building useful databases. Recently, the European Federation of Clinical Chemistry and Laboratory Medicine has set up a new biological variations database called Eubivas [18], which will probably be useful for defining analytical performance goals in the future.

4.3. Which analytical goals should be chosen?

Most of the medical decisions are based on laboratory measurements. In this context, the implementation of quality assurance systems is of particular importance. International norms such as ISO 15189 or ISO 22870 help laboratories in ensuring the quality of results. Quality

assurance requires comparison to performance goals. Considerable efforts have been made in order to provide laboratories with analytical performance goals [14,19]. Multiple authors have provided different analytical performance goals databases, such as the Société Française de Biologie Clinique (SFBC) [14], Ricos et al. [11,12,19], RCPA [13] or CLIA [20]. Linking analytical performance goals to biological variations (BV) [21] is a well-recognized approach in medicine. Goals based on BV, both in physiological and pathological settings are of particular importance because they allow monitoring patients. Considerable work has been done in characterizing BV and deriving performance goals from it. However, data are still lacking and published studies are of unequal quality. Recently, a European Federation of Clinical Chemistry and Laboratory Medicine Task and Finish Group has proposed the Biological Variation Data Critical Appraisal Checklist (BIVAC) [22], in order to improve critical appraisal of published studies of BV. This checklist will help to combine different BV studies sharing high-quality standards and derive robust BV estimates. Our results suggest that in the case of POC implementation, consistency of results across all devices should also be verified. Analysing qualifications results considering all results from all devices as one unique apparatus could represent a way to reach this goal.

For a given parameter the different databases give different goals estimation. Rico's database uses a unique percentage of the analyte at all levels. This approach makes the performance goals independent of analyte level, which is useful. However this approach can be difficult to apply for low levels because a low absolute change represents a large proportion. This limit is compensated when we consider different quality level (optimal, desirable, minimum). On the opposite, SFBC defined unique goals at different levels. Finally, RCPA considered an intermediate approach, providing an absolute error goal for low levels and percentage error above a determined threshold. International consensus regarding the method to apply to establish performance goals would be useful. Efforts are being made to define analytical performance specifications models according to the measurand's role in the decision-making process [23]. Different models have already been proposed, based on outcome (model 1), biological variations (model 2) or state-of-the-art (model 3).

When analytical performance goals derived from BV are set, they have to be compared to analytical performance actually reached by the manufacturers, ie to the state-of-the-art. In this study, 3 devices in the set didn't meet the minimum Ricos' performance for the lowest level of ionized Calcium (0.37 mmol/L). Although such a low level of ionized calcium is not met in physiological conditions, citrate anticoagulation in renal replacement therapy leading to very low ionized calcium levels is widely used nowadays. These results underline the challenges encountered when measuring low levels of ionized calcium using BGA [17]. In some cases, analytical goals should also be considered in view of clinical relevance.

5. Limits and conclusions

Some limits in this study must be acknowledged. In particular, no reference materials were available for all the parameters. Bias and imprecision were thus approached from quality controls provided by Radiometer. Furthermore, goals for serum or plasma were considered when no goals for blood were available. This underlines the need for further study about analytical performance specifications. Finally, linearity and limits of quantitation were not determined in the qualification scheme as they were not mandatory according to the SH GTA 04 document [24].

This study qualified a set of 22 ABL 90 Flex Plus BGA for use for patients care. This study allows evaluation of the ABL 90 Flex Plus analyzer at the unit level and extends these results at a larger scale, that is the whole set of analyzers. Further studies are needed to develop a standardized methodology for qualifying multiple units of analyzers deployed at the same time in multiple locations. This study also stresses the need for further development of analytical performance goals

Acknowledgements

The authors have no conflict of interest to disclose. The authors acknowledge Radiometer for providing quality control materials and reagents for the qualification study. They thank M JP Pouilles for technical assistance.

References

- [1] CLSI, *Blood Gas and pH Analysis and Related Measurements; Approved Guidelines—Second Edition*. CLSI Document C46-A2, Clinical and Laboratory Standards Institute, Wayne, PA, 2009.
- [2] B. Bénétiau-Burnat, P. Pernet, A. Pilon, D. Latour, S. Goujon, A. Feuillu, M. Vaubourdolle, Evaluation of the GEM premier 4000: a compact blood gas CO-Oximeter and electrolyte analyzer for point-of-care and laboratory testing, *Clin. Chem. Lab. Med.* 46 (2) (2008) 271–279, <https://doi.org/10.1515/CCLM.2008.043>.
- [3] M. Oyaert, T. Van Maerken, S. Bridts, S. Van Loon, H. Laverge, V. Stove, Analytical and pre-analytical performance characteristics of a novel cartridge-type blood gas analyzer for point-of-care and laboratory testing, *Clin. Biochem.* 53 (2018) 116–126, <https://doi.org/10.1016/j.clinbiochem.2018.01.007>.
- [4] C. Oris, Y. Clavel, M. Jabaudon, A. Pialat, H.A. Mohamed, ... D. Bouvier, Method validation of a set of 12 GEM® Premier™ 4000 blood gas analyzers for point-of-care testing in a university teaching hospital, *Pract. Lab. Med.* 10 (2018) 21–33, <https://doi.org/10.1016/j.plabm.2017.12.001>.
- [5] A.-S. De Koninck, K. De Decker, J. Van Bocxlaer, P. Meeus, L. Van Hoovels, Analytical performance evaluation of four cartridge-type blood gas analyzers, *Clin. Chem. Lab. Med.* 50 (6) (2012) 1083–1091, <https://doi.org/10.1515/cclm-2011-0685>.
- [6] D.F. Wilson, Oxygen dependent quenching of phosphorescence: a perspective, *Adv. Exp. Med. Biol.* 317 (1992) 195–201, https://doi.org/10.1007/978-1-4615-3428-0_20.
- [7] H.B. Kristensen, Traceability to the Primary Standards At Radiometer Medical. *Radiometer Medical AS117*, (2017).
- [8] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet* 327 (8476) (1986) 307–310, [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8).
- [9] R.A. Parker, C.J. Weir, N. Rubio, R. Rabinovich, H. Pinnock, ... B. McKinstry, Application of mixed effects limits of agreement in the presence of multiple sources of variability: exemplar from the comparison of several devices to measure respiratory rate in COPD patients, *PLoS One* 11 (12) (2016) e0168321, <https://doi.org/10.1371/journal.pone.0168321>.
- [10] CLSI, *Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline—Third Edition*. CLSI Document EP09-A3 Wayne, Clinical and Laboratory Standards Institute, PA, 2013.
- [11] C. Ricós, V. Álvarez, C. Perich, P. Fernández-Calle, J. Minchinela, ... E. González-Lao, Rationale for using data on biological variation, *Clin. Chem. Lab. Med.* 53 (6) (2015) 863–870, <https://doi.org/10.1515/cclm-2014-1142>.
- [12] Desirable Biological Variation Database specifications - Westgard, (n.d.). Retrieved 22 August 2019, from, 2014. <https://www.westgard.com/biodatabase1.htm>.
- [13] RCPA Allowable Limits of Performance for Biochemistry - Westgard, (n.d.). Retrieved 22 August 2019, from, 2012. <https://www.westgard.com/rcpa-biochemistry.htm>.
- [14] A. Vassault, D. Grafmeyer, J. de Graeve, R. Cohen, A. Beaudonnet, J. Bienvenu, Quality specifications and allowable standards for validation of methods used in clinical biochemistry, *Ann. Biol. Clin.* 57 (6) (1999) 685–695.
- [15] A. Marrocco, J.-P. Cristol, A.-M. Boullaran, Setting up of 15 POC blood gas analyzers at Montpellier Hospital (France), *Ann. Biol. Clin.* 74 (1) (2016) 79–92, <https://doi.org/10.1684/abc.2015.1114>.
- [16] M. Uyanik, E. Sertoglu, H. Kayadibi, S. Tapan, M.A. Serdar, ... I. Kurt, Comparison of blood gas, electrolyte and metabolite results measured with two different blood gas analyzers and a core laboratory analyzer, *Scand. J. Clin. Lab. Invest.* 75 (2) (2015) 97–105, <https://doi.org/10.3109/00365513.2014.981854>.
- [17] P. Schwarzer, S.-O. Kuhn, S. Stracke, M. Gründling, S. Knigge, ... A. Petersmann, Discrepant post filter ionized calcium concentrations by common blood gas analyzers in CRRT using regional citrate anticoagulation, *Crit. Care* 19 (1) (2015) 321, <https://doi.org/10.1186/s13054-015-1027-1>.
- [18] A.K. Aarsand, J. Díaz-Garzón, P. Fernandez-Calle, E. Guerra, M. Locatelli, ... European Federation of Clinical Chemistry and Laboratory Medicine Working Group on Biological Variation, The EuBIVAS: within- and between-subject biological variation data for electrolytes, lipids, urea, uric acid, Total protein, Total bilirubin, direct bilirubin, and glucose, *Clin. Chem.* 64 (9) (2018) 1380–1393, <https://doi.org/10.1373/clinchem.2018.288415>.
- [19] C. Ricós, V. Alvarez, F. Cava, J.V. García-Lario, A. Hernández, ... M. Simón, Current databases on biological variation: pros, cons and progress, *Scand. J. Clin. Lab. Invest.* 59 (7) (1999) 491–500.
- [20] CLIA Requirements for Analytical Quality - Westgard, (n.d.). Retrieved 22 August 2019, from, 1992. <https://www.westgard.com/clia.htm>.
- [21] F.C.G. Fraser, *Biological Variation: From Principles To Practice*, AACC Press, Washington, 2001 2001.
- [22] A.K. Aarsand, T. Røraas, P. Fernandez-Calle, C. Ricos, J. Díaz-Garzón, ... S. Sandberg, The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation, *Clin. Chem.* 64 (3) (2018) 501–514, <https://doi.org/10.1373/clinchem.2017.281808>.
- [23] F. Ceriotti, P. Fernandez-Calle, G.G. Klee, G. Nordin, S. Sandberg, ... M. Panteghini, Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM strategic conference, *Clin. Chem. Lab. Med. (CCLM)* 55 (2) (2016) 189–194, <https://doi.org/10.1515/cclm-2016-0091>.
- [24] Comité Français d'accréditation. SH GTA 04, Guide technique d'accréditation de vérification, Retrieved 29 August 2019 from, 2015. <https://tools.cofrac.fr/documentation/SH-GTA-04>.