



# Periodic Comparability Verification and Within-Laboratory Harmonization of Clinical Chemistry Laboratory Results at a Large Healthcare Center With Multiple Instruments

Youngwon Nam , M.D.<sup>1,2</sup>, Joon Hee Lee , M.D.<sup>1,2</sup>, Sung Min Kim , M.T.<sup>2</sup>, Sun-Hee Jun , M.T.<sup>2</sup>, Sang Hoon Song , M.D., Ph.D.<sup>1,3</sup>, Kyunghoon Lee , M.D.<sup>1,2</sup>, and Junghan Song , M.D., Ph.D.<sup>1,2</sup>

<sup>1</sup>Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul, Korea; <sup>2</sup>Department of Laboratory Medicine, Seoul National University Bundang Hospital, Seongnam, Korea; <sup>3</sup>Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea

**Background:** Results from laboratories using multiple instruments should be standardized or harmonized and comparability-verified for consistent quality control. We developed a simple frequent comparability verification methodology applicable to large healthcare centers using multiple clinical chemistry instruments from different manufacturers.

**Methods:** Comparability of five clinical chemistry instruments (Beckman Coulter AU5800, Abbott Architect Ci16000, two Siemens Vista 1500, and Ortho Vitros 5600) was evaluated from 2015 to 2019 for 12 clinical chemistry measurements. Pooled residual patient samples were used for weekly verifications. Results from any instrument exceeding the allowable verification range versus the results from the comparative instrument (AU5800) were reported to clinicians after being multiplied by conversion factors that were determined via a linear regression equation obtained from simplified comparison.

**Results:** Over the five-year study period, 432 weekly inter-instrument comparability verification results were obtained. Approximately 58% of results were converted due to non-comparable verification. Expected average absolute percent bias and percentage of non-comparable results for non-converted and converted results after conversion action were much lower than those for data measured before conversion action. The inter-instrument CV for both non-converted and converted results after conversion action was much lower than that for measured data before conversion action for all analytes.

**Conclusions:** We maintained within-laboratory comparability of clinical chemistry tests from multiple instruments for five years using frequent low-labor periodic comparability verification methods from pooled residual sera. This methodology is applicable to large testing facilities using multiple instruments.

**Key Words:** Comparability, Clinical chemistry, Instruments, Verification

**Received:** October 20, 2020

**Revision received:** April 21, 2021

**Accepted:** September 9, 2021

**Corresponding author:**

Junghan Song, M.D., Ph.D.  
Department of Laboratory Medicine, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea  
Tel: +82-31-787-7691  
Fax: +82-31-787-4015  
E-mail: songjhcp@snu.ac.kr

**Co-corresponding author:**

Kyunghoon Lee, M.D.  
Department of Laboratory Medicine, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea  
Tel: +82-31-787-7696  
Fax: +82-31-787-4015  
E-mail: khlee59023@gmail.com



**© Korean Society for Laboratory Medicine**

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Standardization is the process of obtaining equivalent clinical results from different measurement procedures using a calibration method that is traceable to a reference measurement pro-

cedure (RMP) or certified reference materials (CRMs) [1]. Owing to issues concerning cost, availability, and commutability of RMPs and CRMs, standardization cannot always be achieved in real-world practice. As an alternative to standardization, harmonization encompasses the concept of achieving equivalent re-

sults without a CRM and/or RMP, which implies that results are comparable irrespective of the time and place [2]. Harmonization is used for numerous clinical laboratory measurement procedures that do not have a technically reasonable or commutable reference material and can be a practical solution for calibration traceability for various measurands that would otherwise be non-harmonized [3]. Standardization and/or harmonization ensures the comparability and interchangeability of laboratory results and prevents physicians' confusion and misinterpretation of laboratory results [1, 3, 4].

In large healthcare centers with multiple instruments, patient samples may be exposed to various laboratory test settings. If the instruments for a particular test are not harmonized, different measurement systems may yield biased results for the same sample, resulting in an increased risk of misinterpretation or malpractice. Therefore, periodic comparability verification is necessary to provide clear and harmonized results from different instruments, regardless of the approaches used [5]. However, global standardization or harmonization of clinical laboratory results is currently limited by cost and labor-effectiveness, as well as various technical and environmental challenges, and Korea is no exception. Step-wise protocols for standardization have been developed till date for only a few clinical tests, including cholesterol, creatinine, hormones, and glycohemoglobin [6–10]. More economic and simpler within-laboratory harmonization or comparability verification methods are required to provide convincing and reliable outputs from large laboratories using multiple instruments.

Comparability of results assures that the examination of a measurand is consistent within a laboratory system, even if different methods and instruments are used; however, there is currently no consensus on how to demonstrate such comparability. Laboratories should establish their own procedures to ensure results comparability within their individual systems. The clinical and laboratory standard institute (CLSI) provides guidelines (EP31-A-IR) for comparability verification within a single healthcare system; however, it does not suggest universally accepted criteria or specific methodology for each laboratory [5]. The CLSI guidelines present only limited examples of various situations and methodologies, for which the solutions involve infrequent and complicated processes that are unsuitable for application in specific routine measurements [5, 11]. We propose an easy and frequent periodic comparability verification method that is applicable to various clinical chemistry instruments for the harmonization of diverse tests performed at healthcare centers.

## MATERIALS AND METHODS

### Instruments and test items

Weekly comparability verification of various clinical chemistry tests using multiple instruments was prospectively performed at Seoul National University Bundang Hospital from January 2015 to June 2019. Four different types of instruments were used: Beckman Coulter AU5800 Chemistry Analyzer (Beckman Coulter, Inc., Brea, CA, USA), Abbott Architect Ci16000 Integrated System (Abbott, Abbott Park, IL, USA), two Siemens Dimension Vista 1500 Intelligent Lab systems (Vista1 and Vista2, Siemens Healthineers, Erlangen, Germany), and Ortho Vitros 5600 Integrated System (Vitros, Ortho Clinical Diagnostics, Raritan, NJ, USA). AU5800 was assigned as the standard comparative instrument as this is one of the main instruments of the laboratory used for largest number of clinical chemistry tests and is also a closed system that uses exclusive calibrators and reagents for clinical tests. Reference range setting and proficiency testing (PT), including accuracy-based PT performed by Korean External Quality Assessment Service (KEQAS) [12] and College of American Pathologists (CAP) [13], were also carried out on AU5800.

Three categories of clinical chemistry tests were evaluated for harmonization: (1) electrolytes, including sodium (Na), chloride (Cl), potassium (K), phosphate (P), and calcium (Ca); (2) liver panel (LP), including aspartate transaminase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), albumin, and total protein (TP); and (3) selected standardized test items, including total cholesterol and creatinine. Detailed information regarding the calibrators, reagents, average number of tests per year, imprecision during the study period, and comparative systems for each test item and instrument are shown in Supplemental Data Table S1. For the standardized test items, including total cholesterol and creatinine, commutable reference materials were produced and stored for a nationwide accuracy-based PT.

### Sample preparation for comparability assessment

Residual serum samples were used to prepare sample pools from two to five patients for each biological material considered for comparability assessment. Residual samples were stored for seven days, which were used only during the storage period. Samples were pooled as larger volumes to enable adequate comparability verification between the five instruments. The pooled serum samples originated from residual samples of individuals who visited the hospital or outpatient clinic for healthcare check-ups with venous blood collected for various clinical

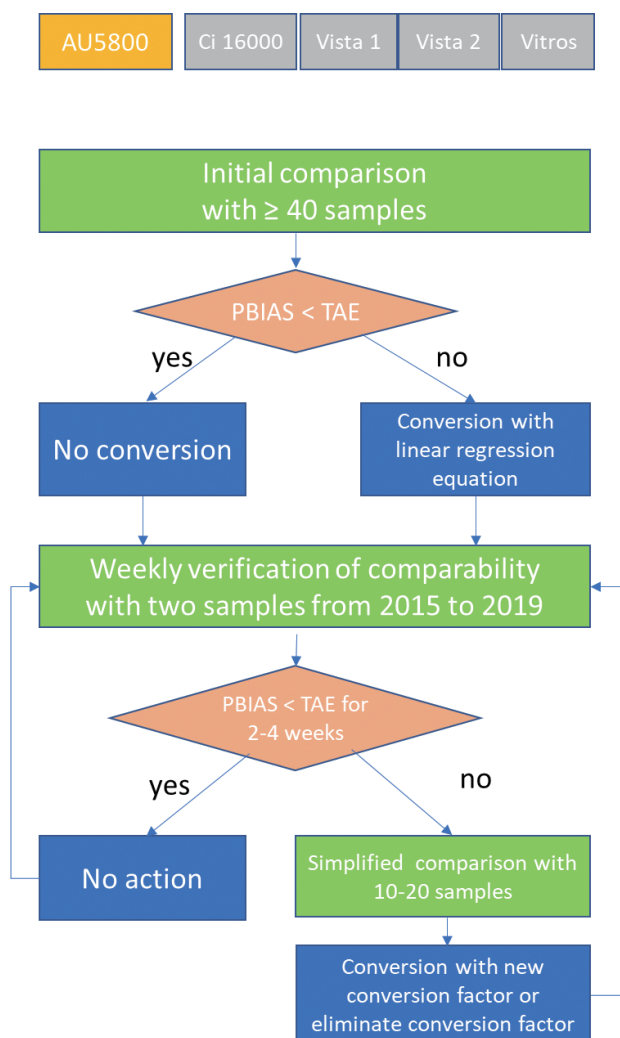
tests in seven days. More than 40 samples for the initial comparison and 10–20 samples for the simplified comparison were prepared at different analyte concentrations spanning the lower and upper limits of the measurement range in accordance with the CLSI EP09-A3 guideline [12].

### Comparability assessment protocol

A schematic diagram of the process used for verification of comparability, including initial comparison, weekly comparability verification, and simplified comparison, across the five different instruments is shown in Fig. 1.

#### Initial comparison and conversion action

The initial comparison was performed according to the CLSI



**Fig. 1.** Schematic diagram of the comparability verification process of five different instruments.  
Abbreviations: PBIAS, percent bias; TAE, total allowable error.

guidelines using more than 40 residual samples [5]. The percent bias (PBIAS) between the comparative AU5800 system and each instrument was calculated. For electrolytes and LP, if the PBIAS for any instrument exceeded the comparability testing acceptance criteria, the results were converted using a normal linear regression equation (intercept,  $a$ ; slope,  $b$ ) to make them comparable to the AU5800 results. The relationship between the measured concentration ( $C_{\text{measured}}$ ) and converted concentration ( $C_{\text{converted}}$ ) was as follows:

$$C_{\text{converted}} = (C_{\text{measured}} - a)/b.$$

To standardize the results of all instruments to the international reference system, an additional comparison was performed for cholesterol and creatinine using CRMs with target values measured using RMPs. When non-comparability existed, the results were converted to be comparable to the reference targets.

#### Weekly comparability verification

Comparability verification of the five instruments was performed each week by measuring LP, electrolytes, cholesterol, and creatinine levels in two pooled serum samples, and comparing the results. To evaluate the harmonized outcome of periodic comparability verification, we calculated the PBIAS of the weekly results between AU5800 and the other instruments and examined the trendline. If the PBIAS exceeded the acceptance criteria of agreement or comparability for two to four weeks, the following simplified comparison was performed.

#### Simplified comparison

If non-comparable results were observed for a specific test or instrument for a few weeks, a simplified comparison was performed using 10–20 of such non-comparable samples for electrolytes and LP or using commutable reference materials for cholesterol and creatinine to determine an appropriate conversion factor according to the linear regression equation. The converted and harmonized results from all instruments following periodic verification were reported to the clinicians rather than the original values. If subsequent results again failed to meet the acceptance criteria, a decision was made to either eliminate the conversion factor or change it to another value based on a new comparison evaluation.

#### Categorization of the results and statistical analysis

All results were categorized into three groups according to each analyte and instrument: (1) measured data without a conversion

action for comparable results (non-converted results), (2) measured data before a conversion action for non-comparable results (measured data before the conversion action), and (3) converted results after the conversion action for non-comparable results (converted results after the conversion action). The number, average absolute PBIAS relative to the reference system, and percentage of non-comparable results (PNR) were analyzed for each category. Inter-instrument variation, expressed as the CV, was also calculated for each analyte according to the three groups before and after the conversion action. Significant differences between groups was evaluated according to the *P*-value based on the paired Student *t* test. Microsoft Excel 16.0 (Microsoft Corporation, Redmond, WA, USA) was used for data recording, comparability checking, and statistical analysis. MedCalc 19.1.7 (MedCalc Software Ltd, Ostend, Belgium) was used for all other analyses.

### Weekly comparability verification and simplified comparison

Over the five years of the study period, approximately 432 weekly inter-instrument comparability verification results were collected for the three categories (electrolytes, LP, and cholesterol/creatinine) in pooled residual serum samples. We applied the Royal College of Pathologists of Australasia (RCPA) total allowable limit of performance goals to our database and then compared the weekly results from each instrument to determine whether they were within the allowable range compared with the AU5800 results (Table 1) [13]. Simplified comparison was performed for data that exceeded the allowable range of comparability (non-comparability data). For the within-laboratory harmonization of the clinical chemistry tests, converted results were reported to clinicians after the measured results were multiplied by the appropriate conversion factor obtained via linear regression analysis of the simplified comparison results (conversion action). The history of the conversion action is outlined in Supplemental Data Table S2. This research was granted review exemption by the Seoul National University Bundang Hospital Institutional Review Board (IRB No. X-2108-705-903).

## RESULTS

### Initial comparison

Following analysis of the initial comparison of more than 40 samples, initial conversion factors were applied to the instruments for specific items when the PBIAS between the comparative system and individual instruments was greater than the agreement acceptance criteria. Initial conversion factors were

**Table 1.** Allowable limits of performance in the RCPAQAP general serum chemistry programs\*

Analyte	Unit	Fixed deviation	To	TAE	Harmonization medical impact/status
AST	U/L	5	40	12%	Medium/Incomplete
ALT	U/L	5	40	12%	Medium/Incomplete
ALP	U/L	15	125	12%	Medium/Incomplete
ALB	g/dL	0.2	3.3	6%	Medium/Needed
TP	g/dL	0.3	6	5%	Medium/Incomplete
Na	mmol/L	3	150	2%	High/Adequate, Maintain
K	mmol/L	0.2	4.0	5%	High/Adequate, Maintain
Cl	mmol/L	3	100	3%	High/Adequate, Maintain
Ca	mg/dL	0.40	10.0	4%	High/Adequate, Maintain
P	mg/dL	0.57	7.12	8%	Medium/Adequate, Maintain
Chol	mg/dL	11.6	193.3	6%	High/Adequate, Maintain
Cr	mg/dL	0.09	1.13	8%	High/Adequate, Maintain

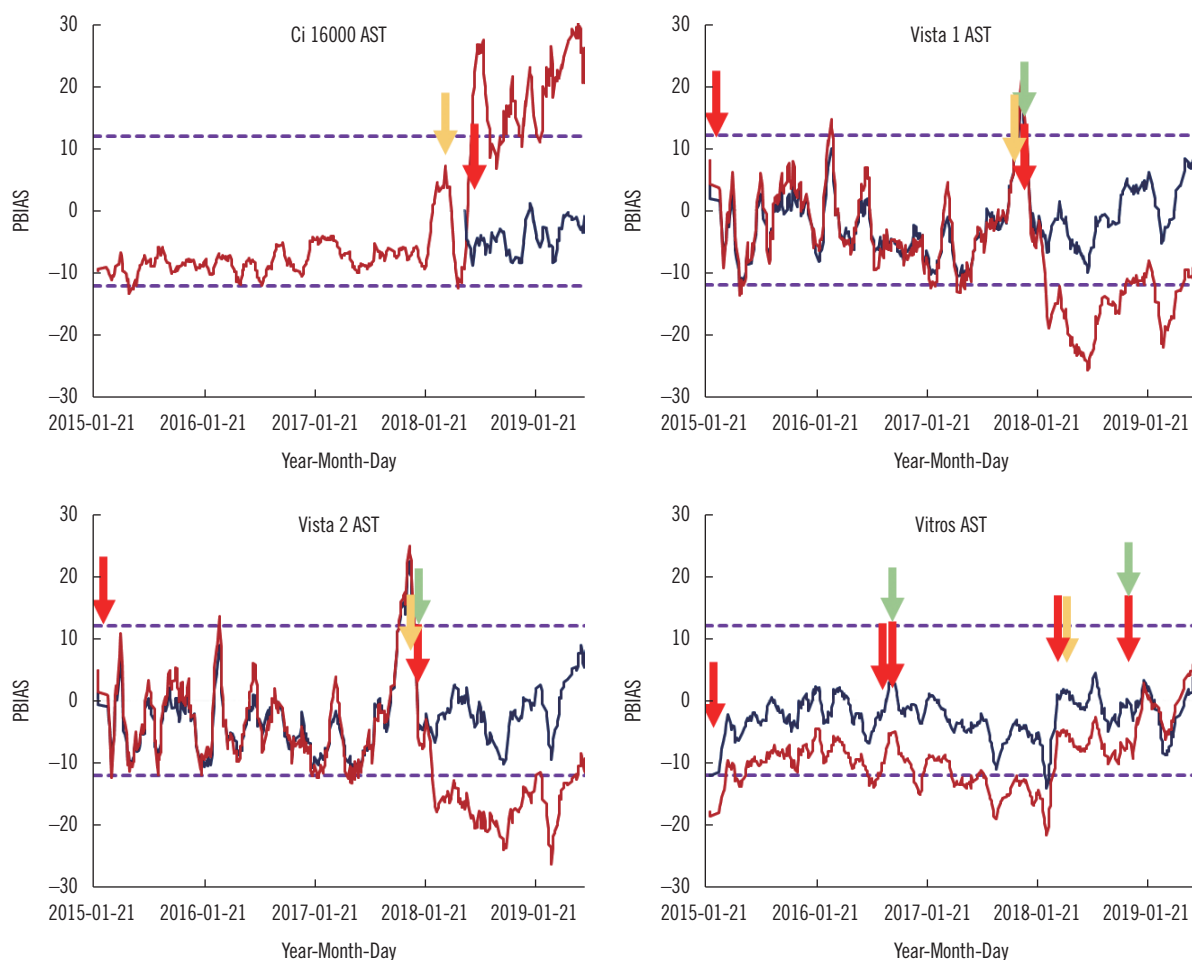
\*These parameters were used as the acceptance criteria and allowable limits for the within-laboratory comparability verification of multiple instruments. The allowable limits have fixed deviations (+/-) from the target or consensus values up to a particular value (To) and proportional deviations (TAE) at higher values.

Abbreviations: AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; ALB, albumin; TP, total protein; Na, sodium; K, potassium; Cl, chloride; Ca, calcium; P, phosphate; Chol, cholesterol; Cr, creatinine; TAE, total allowable error; RCPAQAP, Royal College of Pathologists of Australasia Quality Assurance Program.

applied to 62% (31/50) of the test results from the various instruments, as shown in column “2015-02-03” of Supplemental Data Table S2.

### Trendlines of weekly comparability verification

Fig. 2 shows examples of PBIAS trendlines for AST. Among the three test categories, verification of LP showed the most obvious changes from unconverted results, which may lead to non-comparability due to significant differences in calibrators or reagents. For example, converted AST results from multiple instruments showed good within-laboratory comparability until the summer of 2017 (in allowable limits,  $\pm 12\%$ ), whereas the original data before conversion action showed unacceptable PBIAS for several instruments, especially for the Vitros instrument in 2017 ( $-28.5\%$ ) and for the Vista instrument in 2017 ( $-19.2\%$ ) and 2018 ( $-22.7\%$ ). When the reagent was switched from one not containing pyridoxal-5-phosphate (P5P) to one containing P5P in March 2018 in all instruments, the original PBIAS of the Vista instrument was reduced to  $-20\%$ . According to this PBIAS drop for AST tests, we calculated new conversion factors for each instrument to adjust the within-laboratory harmonization of



**Fig. 2.** PBIAS trendlines of AST for the Ci16000, two Vista, and Vitros instruments. Blue lines are converted PBIAS trendlines, and red lines are original (unconverted) PBIAS trendlines. Calibrator changes are indicated with a green arrow, and conversion factor changes are indicated with a red arrow. Reagent changes from without P5P to with P5P are indicated with a yellow arrow. RCPA acceptance criteria for AST comparability total error is under 12%, as depicted by the two purple dotted lines [15]. Abbreviations: AST, aspartate transaminase; PBIAS, percent bias; RCPA, Royal College of Pathologists of Australasia.

AST results, and comparability was successfully achieved (allowable  $\text{PBIAS} \pm 10\%$ ). As the original results for the Vitros instrument had become non-comparable with the reference instrument due to the P5P-containing reagent, the conversion factor was again modified to harmonize the AU5800 and Vitros results and maintain comparability.

Although the acceptable PBIAS ranges were narrower in the electrolyte category relative to those in the other clinical chemistry test categories, the converted results from all instruments reported to clinicians were in accordance with the acceptance criteria and comparable to the reference instrument; however, the original results for several tests from the Vista and Vitros instruments were outside the acceptable range and non-comparable to the AU5800 results. Serum calcium measurement from the

Vitros instrument originally showed high PBIAS relative to the reference instrument (2%–10%), and the converted PBIAS oscillated within the acceptable range (–2% to +4%).

Unlike most tests that maintained within-laboratory harmonization and comparability with the AU5800 reference system, tests with an available nationwide accuracy-based PT program, including serum creatinine and cholesterol, showed slightly different patterns of comparability verification compared with other categories. Therefore, creatinine and cholesterol results were adjusted to the international reference values by recalculating conversion factors if non-comparable results were obtained.

#### Distribution of absolute PBIAS and PNR

The average absolute PBIAS (12.8%) and PNR (23.8% of mea-

**Table 2.** Number of data points, average absolute PBIAS, and PNR for each test item and instrument according to three data categories

Analyte	Instrument	Measured data without conversion action*			Data before (measured) <sup>†</sup> and after (converted) conversion action				
		N (%) <sup>‡</sup>	Average absolute PBIAS (95% CI) <sup>§</sup>	PNR <sup>  </sup> (% PNR)	N (%)	Measured results		Converted results	
						Average absolute PBIAS (95% CI) <sup>§</sup>	PNR (% PNR)	Average absolute PBIAS (95% CI)	PNR (% PNR)
AST	Ci16000 <sup>¶</sup>	338 (78)	8.1 (7.5–8.6)	1.5 (1.2)	94 (22)	49.6 (48.3–51.0)	64.9 (14.1)	7.1 (5.6–8.6)	3.2 (0.7)
	Vista1 <sup>¶</sup>	0 (0)			432 (100)	11.6 (10.4–12.8)	25.5 (25.5)	7.2 (6.4–8.1)	3.9 (3.9)
	Vista2 <sup>¶</sup>	0 (0)			432 (100)	12.1 (10.8–13.3)	26.6 (26.6)	7.6 (6.7–8.5)	5.1 (5.1)
	Vitros <sup>¶</sup>	2 (0)	12.5	0.0 (0.0)	430 (100)	9.9 (9.3–10.5)	19.5 (19.4)	4.9 (4.3–5.4)	0.9 (0.9)
ALT	Ci16000 <sup>¶</sup>	401 (93)	5.2 (4.7–5.8)	0.0 (0.0)	31 (7)	104.3	67.7 (4.9)	23.4	29.0 (2.1)
	Vista1 <sup>¶</sup>	0 (0)			432 (100)	31.6 (29.9–33.4)	75.5 (75.5)	9.7 (8.5–10.8)	2.3 (2.3)
	Vista2 <sup>¶</sup>	0 (0)			432 (100)	30.8 (29.2–32.4)	72.5 (72.5)	9.3 (8.2–10.4)	2.5 (2.5)
	Vitros <sup>¶</sup>	2 (0)	9.3	0.0 (0.0)	430 (100)	14.7 (13.4–16.0)	23.7 (23.6)	9.0 (7.9–10.1)	1.4 (1.4)
ALP	Ci16000 <sup>¶</sup>	98 (23)	5.1 (4.4–5.7)	0.0 (0.0)	334 (77)	12.5 (11.6–13.4)	22.8 (17.6)	4.0 (3.5–4.5)	0.0 (0.0)
	Vista1 <sup>¶</sup>	0 (0)			432 (100)	4.5 (4.1–4.9)	1.2 (1.2)	4.1 (3.6–4.5)	0.0 (0.0)
	Vista2 <sup>¶</sup>	0 (0)			432 (100)	4.6 (4.3–5.0)	1.9 (1.9)	4.2 (3.7–4.7)	0.7 (0.7)
	Vitros <sup>¶</sup>	0 (0)			432 (100)	9.2 (8.5–9.9)	24.1 (24.1)	7.2 (6.5–7.9)	5.1 (5.1)
ALB	Ci16000 <sup>¶</sup>	372 (86)	2.3 (2.0–2.6)	3.5 (3.0)	60 (14)	16.5 (16.2–16.7)	11.7 (1.6)	4.7 (3.9–5.5)	23.3 (3.2)
	Vista1 <sup>¶</sup>	18 (4)	4.6	27.8 (1.2)	414 (96)	7.3 (6.9–7.6)	58.9 (56.5)	3.5 (3.2–3.8)	12.8 (12.3)
	Vista2 <sup>¶</sup>	18 (4)	5.7	44.4 (1.9)	414 (96)	7.6 (7.2–7.9)	60.4 (57.9)	3.3 (3.0–3.6)	11.4 (10.9)
	Vitros <sup>¶</sup>	169 (39)	3.0 (2.4–3.5)	6.5 (2.5)	263 (61)	5.3 (4.9–5.7)	17.1 (10.4)	3.3 (2.9–3.7)	8.0 (4.9)
TP	Ci16000 <sup>¶</sup>	372 (86)	1.4 (1.2–1.5)	0.0 (0.0)	60 (14)	9.7 (9.5–9.8)	0.0 (0.0)	1.4 (1.0–1.8)	0.0 (0.0)
	Vista1 <sup>¶</sup>	292 (68)	1.8 (1.6–2.1)	2.7 (1.9)	140 (32)	6.3 (6.1–6.5)	12.9 (4.2)	3.0 (2.6–3.3)	12.9 (4.2)
	Vista2 <sup>¶</sup>	292 (68)	1.5 (1.2–1.7)	1.4 (0.9)	140 (32)	5.3 (5.1–5.5)	7.9 (2.5)	3.1 (2.7–3.4)	8.6 (2.8)
	Vitros <sup>¶</sup>	290 (67)	1.6 (1.4–1.8)	2.4 (1.6)	142 (33)	5.4 (5.2–5.6)	7.7 (2.5)	2.5 (2.1–2.9)	4.9 (1.6)
Na	Ci16000	432 (100)	1.0 (0.9–1.1)	0.9 (0.9)	0 (0)				
	Vista1 <sup>¶</sup>	294 (68)	1.0 (0.9–1.1)	3.4 (2.3)	138 (32)	2.9 (2.8–3.0)	2.9 (0.9)	1.1 (0.9–1.2)	3.6 (1.2)
	Vista2 <sup>¶</sup>	294 (68)	1.0 (0.9–1.2)	5.4 (3.7)	138 (32)	3.1 (3.0–3.2)	9.4 (3.0)	1.2 (1.0–1.4)	5.1 (1.6)
	Vitros <sup>¶</sup>	398 (92)	1.1 (1.0–1.3)	6.5 (6.0)	34 (8)	15.7	55.9 (4.4)	1.0	2.9 (0.2)
K	Ci16000	432 (100)	1.1 (0.9–1.2)	0.2 (0.2)	0 (0)				
	Vista1	432 (100)	2.0 (1.9–2.2)	0.0 (0.0)	0 (0)				
	Vista2	432 (100)	2.0 (1.9–2.2)	0.7 (0.7)	0 (0)				
	Vitros <sup>¶</sup>	0 (0)			432 (100)	2.1 (1.9–2.3)	3.7 (3.7)	1.4 (1.2–1.6)	0.5 (0.5)
Cl	Ci16000	432 (100)	0.9 (0.8–1.0)	0.7 (0.7)	0 (0)				
	Vista1 <sup>¶</sup>	4 (1)	3.8	75.0 (0.7)	428 (99)	2.3 (2.2–2.5)	28.0 (27.8)	1.2 (1.0–1.3)	0.9 (0.9)
	Vista2 <sup>¶</sup>	4 (1)	2.4	0.0 (0.0)	428 (99)	2.1 (1.9–2.2)	21.0 (20.8)	1.3 (1.1–1.4)	2.6 (2.5)
	Vitros <sup>¶</sup>	10 (3)	12.5	20.0 (0.6)	306 (97)	1.3 (1.1–1.4)	3.6 (3.5)	1.2 (1.0–1.4)	2.6 (2.5)
Ca	Ci16000	198 (46)	1.6 (1.3–1.8)	2.0 (0.9)	234 (54)	3.6 (3.4–3.8)	10.3 (5.6)	1.8 (1.5–2.0)	2.6 (1.4)
	Vista1 <sup>¶</sup>	0 (0)			432 (100)	4.1 (3.8–4.4)	45.8 (45.8)	1.9 (1.7–2.1)	5.6 (5.6)
	Vista2 <sup>¶</sup>	0 (0)			432 (100)	4.3 (4.0–4.6)	48.4 (48.4)	2.0 (1.8–2.2)	6.3 (6.3)
	Vitros <sup>¶</sup>	0 (0)			432 (100)	5.7 (5.5–6.0)	66.9 (66.9)	2.1 (1.8–2.3)	8.6 (8.6)
P	Ci16000	432 (100)	3.9 (3.6–4.2)	0.0 (0.0)	0 (0)				
	Vista1 <sup>¶</sup>	0 (0)			432 (100)	2.9 (2.5–3.2)	0.0 (0.0)	4.4 (4.1–4.7)	0.0 (0.0)
	Vista2 <sup>¶</sup>	0 (0)			432 (100)	3.0 (2.7–3.4)	0.0 (0.0)	4.4 (4.0–4.7)	0.2 (0.2)
	Vitros <sup>¶</sup>	0 (0)			432 (100)	10.0 (9.5–10.5)	11.3 (11.3)	4.6 (4.1–5.0)	0.2 (0.2)

(Continued to the next page)

Table 2. Continued

Analyte	Instrument	Measured data without conversion action*			Data before (measured) <sup>†</sup> and after (converted) conversion action				
		N (%) <sup>‡</sup>	Average absolute PBIAS (95% CI) <sup>§</sup>	PNR <sup>  </sup> (% PNR)	N (%)	Measured results		Converted results	
					Average absolute PBIAS (95% CI) <sup>§</sup>	PNR (% PNR)	Average absolute PBIAS (95% CI)	PNR (% PNR)	
Chol	Ci16000 <sup>¶</sup>	0 (0)			432 (100)	2.3 (2.0–2.6)	0.0 (0.0)	2.2 (2.0–2.5)	0.0 (0.0)
	Vista1 <sup>¶</sup>	378 (88)	3.0 (2.7–3.4)	0.8 (0.7)	54 (13)	24.4 (24.0–24.7)	3.7 (0.5)	4.1 (3.2–5.1)	0.0 (0.0)
	Vista2 <sup>¶</sup>	378 (88)	2.5 (2.2–2.8)	0.0 (0.0)	54 (13)	20.6 (20.3–20.9)	1.9 (0.2)	3.5 (2.7–4.3)	0.0 (0.0)
	Vitros <sup>¶</sup>	374 (87)	2.4 (2.1–2.7)	0.0 (0.0)	58 (13)	18.8 (18.5–19.1)	0.0 (0.0)	3.0 (2.0–3.9)	0.0 (0.0)
Cr	Ci16000 <sup>¶</sup>	0 (0)			432 (100)	5.9 (5.3–6.5)	15.8 (15.8)	4.2 (3.7–4.6)	2.5 (2.8)
	Vista1	296 (69)	4.1 (3.6–4.7)	5.7 (3.9)	136 (31)	16.0 (15.4–16.6)	27.2 (8.6)	4.9 (3.9–5.9)	5.1 (1.6)
	Vista2	296 (69)	3.9 (3.4–4.4)	1.7 (1.2)	136 (31)	14.8 (14.3–15.4)	19.1 (6.0)	4.2 (3.4–5.1)	3.7 (1.2)
	Vitros	432 (100)	5.1 (4.6–5.5)	8.1 (8.1)	0 (0)				
Average <sup>¶</sup>	179 (42)	3.7	6.9 (1.4)	250 (58)	12.8	23.8 (17.5)	4.3	4.6 (2.5)	

\*Results that showed acceptable comparability to the reference instrument without a converting action. <sup>†</sup>Results with repeated unacceptable comparability to the reference instrument before conversion, but more acceptable comparability after conversion. <sup>‡</sup>The number of times the comparability tests were performed over five years with percentage relative to all available tests in parentheses. <sup>§</sup>95% confidence intervals were calculated only when the number of cases was more than 50. <sup>||</sup>The percentage of unacceptable, incomparable data outside the total allowable limits relative to the number of data points. The percentage relative to total tests (N=432) is shown in parentheses, which can be calculated by multiplying by the percentage of each number of times to total available tests. <sup>¶</sup>Statistically significant difference between results before and after the conversion action ( $P < 0.0001$ ).

Abbreviations: PBIAS, percent bias; AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; ALB, albumin; TP, total protein; Na, sodium; K, potassium; Cl, chloride; Ca, calcium; P, phosphate; Chol, cholesterol; Cr, creatinine; PNR, percentage of noncomparable results.

sured data with conversion action, 17.5% of total results) from data before the conversion action were significantly higher than those from data after conversion action (4.3%,  $P = 0.0005$  and 4.6%, 2.5% of total results,  $P < 0.0001$ , respectively) (Table 2). This tendency was observed for all instruments and test items, with only a few exceptions when group numbers were small.

Among the 432 comparability data points for each test item and instrument, approximately 58% were converted due to non-comparable results. The rates of conversion action differed according to the test item and instrument. In general, the rates of conversion of LP and electrolytes were lower for Ci16000 (27% and 11%, respectively) and higher for Vista (86% and 66%, respectively) and Vitros (79% and 81%, respectively). The rates of conversion action for cholesterol and creatinine were higher for Ci16000 (100% each) and lower for Vista (13% and 31%, respectively) and Vitros (13% and 0%, respectively).

### Inter-instrument CV before and after conversion action

For all analytes, the CVs for both measured data without conversion action and converted results after conversion action were much lower than those for the measured data before conversion action (Table 3). The inter-instrument CVs of the converted results after conversion action were the lowest for electrolytes and higher for LP. We compared the ratio of inter-instrument CV to

Table 3. Inter-instrument CV (%) and ratio to allowable bias

Analytes	Measured results without conversion action	Original results before conversion action	Converted results after conversion action	Allowable bias*
AST	4.3 (0.8)	8.6 (1.7)	4.9 (1.0)	5.1
ALT	3.2 (0.3)	17.0 (1.5)	7.4 (0.7)	11.1
ALP	3.5 (0.8)	6.0 (1.4)	4.0 (0.9)	4.3
ALB	2.0 (0.6)	4.0 (1.3)	2.4 (0.8)	3.2
TP	1.3 (0.7)	1.5 (0.8)	1.0 (0.5)	2.0
Na	0.8 (1.0)	1.2 (1.5)	0.7 (0.9)	0.8
K	1.3 (0.7)	1.3 (0.7)	0.8 (0.4)	1.9
Cl	0.6 (0.4)	1.3 (0.9)	0.9 (0.6)	1.4
Ca	1.0 (0.5)	3.6 (1.8)	1.5 (0.8)	2.0
P	2.1 (0.8)	4.6 (1.6)	3.0 (1.1)	2.8
Cr	All converted	5.8 (1.3)	4.5 (1.0)	4.4
Chol	All converted	2.1 (0.6)	1.7 (0.5)	3.4

\*Allowable limits for bias according to European recommendations adopted from Baadenhuijsen, *et al.* [16].

Abbreviations: AST, aspartate transaminase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; ALB, albumin; TP, total protein; Na, sodium; K, potassium; Cl, chloride; Ca, calcium; P, phosphate; Chol, cholesterol; Cr, creatinine.

the allowable bias by European recommendations for each category [16]. Most ratios from the measured data without conversion action and converted results after the conversion action

were  $<1$ . In contrast, ratios from original data before conversion action were  $>1$ , suggesting that the inter-instrument CV from the original data before the conversion action was not within the allowable bias.

## DISCUSSION

The International Consortium for Harmonization of Clinical Laboratory Results (ICHCLR) was established to fulfill the recommendations of an international conference in 2011 convened to review the available infrastructure and challenges in achieving harmonization of results among different measurement procedures [3]. To achieve simple within-laboratory harmonization of clinical tests in a healthcare center, the CLSI EP31-A-IR guidelines suggest periodic comparability verification that can be applied to various circumstances, even in the absence of universally accepted criteria or specific methodology [5]. These guidelines suggest practical and general considerations, including which samples can be used for verification, frequency of assessment protocols, and examples of various acceptance criteria for comparability testing. However, the example regarding comparability verification of clinical chemistry tests in the appendix of the guidelines would require the preparation of specific replicate samples for multiple runs of each test on each device and specific acceptance criteria for each category. This methodology would likely be labor-intensive and expensive, making it difficult to follow frequently and only suitable for laboratories with few instruments from the same manufacturer [5, 11]. We developed a simple and frequent comparability verification methodology applicable to large healthcare centers with multiple clinical chemistry measurement instruments from different manufacturers.

In our laboratory, clinical chemistry tests were performed using five instruments from four manufacturers. To perform both cost-effective and labor-effective periodic comparability verification, we prepared pooled human sera from residual samples. Each comparability verification was performed weekly using the pooled sera and then by determining whether the results compared with the reference instrument (AU5800) were within the total allowable error (TAE) limits based on the RCPAQAP [15]. Although the comparability of our five instruments was monitored at relatively short intervals, this approach did not incur any additional costs other than limited labor and time for sample collection and pooling. The verified and corrected results from multiple instruments showed good comparability over the five years except for a few minor incidents that were promptly ad-

justed once non-comparability was detected. Therefore, we recommend this simple comparability verification methodology for laboratories in similar environments.

The average absolute PBIAS and PNR were much lower in the converted results than in the original data before conversion action. This enabled reporting more accurate and harmonized results to clinicians, even though we could not report all results within the acceptance criteria of TAEs. If weekly comparability verification had not been performed, our laboratory would have reported approximately 17.5% erroneous and non-harmonized results to clinicians; with its implementation, these errors were reduced to approximately 2.5% depending on the instrument. We found lower inter-instrument CVs in the converted results than in the original data before conversion action. Most CV ratios of the original data to allowable bias were  $>1.0$ , except for TP, potassium, chloride, and cholesterol, whereas the CV ratios of converted results to allowable bias were  $<1.0$ , except for phosphorus and creatinine. These results suggest that our weekly verification of comparability and appropriate conversion action resulted in successful intra-laboratory harmonization of the five instruments.

Serum electrolytes are important factors in many clinical decision-making situations, and their analysis in a healthcare center may be performed serially or periodically using different measuring instruments. Serum electrolyte tests should have a narrow range of acceptance criteria for laboratory comparability. According to the ICHCLR report of measurands, electrolytes such as sodium, potassium, and chloride should be adequately harmonized and maintained given their high medical impact [17]. In our weekly comparability verification, these three electrolytes showed relatively low PBIAS and PNR along with low inter-instrument variation, confirming their excellent harmonization status.

For LP items, which have wider allowable limits for comparability than electrolytes, comparability verification was successfully performed during the five years, except for intermittent events, including changes in reagents containing P5P. As P5P is a coenzyme of the AST and ALT reaction in the human body, it is included in the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) recommendations for measuring aminotransferase levels [18, 19]. However, several laboratories use a modified IFCC method assuming that P5P concentrations in patient serum are sufficient and therefore reagents without P5P can be used for AST and ALT tests [20, 21]. We discourage the modified IFCC method on the basis of its variable total error, which cannot be corrected through calibra-



tion [20]. As serum P5P levels in patients are often insufficient, serum AST and ALT levels tend to be lower when using the modified IFCC method, particularly in patients with poor health status, such as those with acute myocardial infarction [22]. After we changed reagents in February 2018 to meet IFCC recommendations, PBIAS trendlines of AST comparability verification changed remarkably. However, we immediately detected the event and successfully adjusted the comparability of the reported LP via our periodic comparability verification and conversion factor adjustments.

A high level of testing accuracy is clinically important for analytes such as creatinine and hemoglobin A1c (HbA1c). Accuracy-based PT is mandatory for a laboratory to determine the accuracy of patient results by comparing them with reference method results. These survey programs have been established in Korea using commutable reference materials with true target values traceable to CRMs and/or RMPs [13, 23, 24]. When inter-instrument non-comparability was detected, we obtained conversion factors traceable to reference target values using commutable reference materials that had been used for national accuracy-based surveys, instead of results from the reference instrument (AU5800). The inter-instrument CVs of converted results for creatinine and cholesterol were lower than those of LP, but higher than those of electrolytes.

A limitation of our weekly comparability verification was that we arbitrarily assigned the AU5800 instrument as the comparative reference system. Although AU5800 is the main instrument used in our laboratory with a closed system that uses exclusive calibrators and reagents for clinical tests with periodic monitoring of accuracy by performing PT from KEQAS and CAP, it is also susceptible to random or systematic errors, which could affect the comparability to the other instruments. However, we could detect errors in AU5800 indirectly since the other instruments showed collective bias to the reference instrument.

Another limitation of this study was that we could not guarantee the commutability of the samples for weekly comparability testing, even though we tried to pool serum samples without interfering materials. However, we used fully commutable materials for cholesterol and creatinine level measurements in the simplified comparison. The final limitation was that several comparability verification failure events were still observed owing to the fast, unexpected elevation of the PBIAS following changes in the reagent or calibrator. To prevent verification failure events, it is necessary to carefully check comparability results when there are specific events that can greatly affect within-laboratory harmonization. Our simple comparability verification is not a

complete alternative to the CLSI EP31-A-IR guidelines [5], but an adjunct to improve intra-laboratory harmonization and minimize bias.

In conclusion, we succeeded in maintaining within-laboratory comparability of clinical chemistry tests from multiple instruments for five years employing a more frequent (weekly) and less laborious periodic comparability verification method using pooled residual human sera. Our method may assist the existing labor-intensive and complicated method based on the CLSI EP31-A-IR guidelines [5]. This comparability verification method can be applied to large healthcare centers or clinical laboratories that use multiple types of instruments.

## ACKNOWLEDGEMENTS

None.

## AUTHOR CONTRIBUTIONS

Conceptualization: Lee K, Song SH, and Song J. Methodology: Lee K and Song J. Formal analysis: Nam Y, Lee JH, and Kim SM. Data curation: Nam Y, Lee JH, Kim SM, Jun SH, and Lee K. Writing-original draft preparation: Nam Y. Writing-review and editing: Lee JH, Song SH, Lee K, and Song J. Supervision: Lee K, Song SH, and Song J. All authors have accepted responsibility for the entire content of this manuscript and approved the submission.

## CONFLICT OF INTEREST

None declared.

## RESEARCH FUNDING

None declared.

## ORCID

Youngwon Nam	<a href="https://orcid.org/0000-0001-7049-1990">https://orcid.org/0000-0001-7049-1990</a>
Joon Hee Lee	<a href="https://orcid.org/0000-0002-9649-1371">https://orcid.org/0000-0002-9649-1371</a>
Sung Min Kim	<a href="https://orcid.org/0000-0002-8252-1269">https://orcid.org/0000-0002-8252-1269</a>
Sun-Hee Jun	<a href="https://orcid.org/0000-0003-1232-9113">https://orcid.org/0000-0003-1232-9113</a>
Sang Hoon Song	<a href="https://orcid.org/0000-0002-5084-1137">https://orcid.org/0000-0002-5084-1137</a>
Kyunghoon Lee	<a href="https://orcid.org/0000-0002-3154-0347">https://orcid.org/0000-0002-3154-0347</a>
Junghan Song	<a href="https://orcid.org/0000-0003-0576-9938">https://orcid.org/0000-0003-0576-9938</a>

## REFERENCES

1. Miller WG. Harmonization: its time has come. *Clin Chem* 2017;63:1184-6.
2. Miller WG, Tate JR, Barth JH, Jones GR. Harmonization: the sample, the measurement, and the report. *Ann Lab Med* 2014;34:187-97.
3. Greg Miller W, Myers GL, Lou Gantzer M, Kahn SE, Schönbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 2011;57:1108-17.
4. Tate JR and Myers G. Harmonization of clinical laboratory test results. *EJIFCC* 2016;27:5-14.
5. CLSI. Verification of comparability of patient results within one health care system: approved guideline (interim revision). CLSI EP31-A-IR. Wayne, PA: Clinical and Laboratory Standards Institute. 2012.
6. Myers GL, Kimberly MM, Waymack PP, Smith SJ, Cooper GR, Sampson EJ. A reference method laboratory network for cholesterol: A model for standardization and improvement of clinical laboratory measurements. *Clin Chem* 2000;46:1762-72.
7. Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T, et al. Recommendations for improving serum creatinine measurement: A report from the laboratory working group of the national kidney disease education program. *Clin Chem* 2006;52:5-18.
8. Vesper HW and Botelho JC. Standardization of testosterone measurements in humans. *J Steroid Biochem Mol Biol* 2010;121:513-9.
9. Thienpont LM, Van Uytffanghe K, De Grande LAC, Reynders D, Das B, Faix JD, et al. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clin Chem* 2017;63:1248-60.
10. Little RR, Rohlfing C, Sacks DB. The National Glycohemoglobin Standardization Program: over 20 years of improving hemoglobin A1c measurement. *Clin Chem* 2019;65:839-48.
11. Lee EJ, Lee E, Kim M, Kim H-S, Lee YK, Kang HJ. Verification of the comparability of laboratory results from two instruments within one health care system according to Clinical and Laboratory Standard Institute EP31-A-IR. *J Lab Med Qual Assur* 2016;38:129-36.
12. CLSI. Measurement procedure comparison and bias estimation using patient samples; Approved guideline-Third edition. CLSI document EP09-A3. Wayne, PA: Clinical Laboratory Standards Institute, 2013.
13. Kim S, Lee K, Park HD, Lee YW, Chun S, Min WK. Schemes and performance evaluation criteria of Korean Association of External Quality Assessment (KEQAS) for improving laboratory testing. *Ann Lab Med* 2021;41:230-9.
14. College of American Pathologists (CAP). External Quality Assurance/proficiency testing for international laboratories. <https://www.cap.org/laboratory-improvement/international-laboratories/external-quality-assurance-proficiency-testing-for-international-laboratories> (Updated on September 2021).
15. Jones GR, Sikaris K, Gill J. "Allowable limits of performance" for External Quality Assurance programs – an approach to application of the Stockholm criteria by the RCPA Quality Assurance programs. *Clin Biochem Rev* 2012;33:133-9.
16. Baadenhuijsen H, Scholten R, Willems HL, Weykamp CW, Jansen RT. A model for harmonization of routine clinical chemistry results between clinical laboratories. *Ann Clin Biochem* 2000;37:330-7.
17. American Association of Clinical Chemistry (AACC). International consortium for harmonization of clinical laboratory results (ICHCLR). <http://www.harmonization.net/measurands> (Updated on July 2021).
18. Schumann G, Bonora R, Ceriotti F, Féraud G, Ferrero CA, Franck PF, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. International Federation of Clinical Chemistry and Laboratory Medicine. Part 5. Reference procedure for the measurement of catalytic concentration of aspartate aminotransferase. *Clin Chem Lab Med* 2002;40:725-33.
19. Schumann G, Bonora R, Ceriotti F, Féraud G, Ferrero CA, Franck PF, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. International Federation of Clinical Chemistry and Laboratory Medicine. Part 4. Reference procedure for the measurement of catalytic concentration of alanine aminotransferase. *Clin Chem Lab Med* 2002;40:718-24.
20. Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, et al. A category 1 EQA scheme for comparison of laboratory performance and method performance: an international pilot study in the framework of the Calibration 2000 Project. *Clin Chim Acta* 2014;432:90-8.
21. Jun SH and Song J. Annual report on the external quality assessment scheme for clinical chemistry in Korea (2014). *J Lab Med Qual Assur* 2015;37:115-23.
22. Lee JS, Lee K, Kim SM, Choi MS, Jun SH, Song WH, et al. Effects of pyridoxal-5'-phosphate on aminotransferase activity assay. *Lab Med Online* 2017;7:128-34.
23. Jeong TD, Lee HA, Lee K, Yun YM. Accuracy-based proficiency testing of creatinine measurement: 7 years' experience in Korea. *J Lab Med Qual Assur* 2019;41:13-23.
24. Kim JH, Cho Y, Lee SG, Yun YM. Report of Korean association of external quality assessment service on the accuracy-based lipid proficiency testing (2016-2018). *J Lab Med Qual Assur* 2019;41:121-9.