



Immunosuppressive Drug Measurement by Liquid Chromatography Coupled to Tandem Mass Spectrometry: Interlaboratory Comparison in the Korean Clinical Laboratories

Hyun-Ki Kim , M.D.¹, Hyung-Doo Park , M.D.², Sang-Guk Lee , M.D.³, Hyojin Chae , M.D.⁴, Sang Hoon Song , M.D.⁵, Yong-Wha Lee , M.D.⁶, Yeo-Min Yun , M.D.⁷, Sunhyun Ahn , M.D.⁸, Serim Kim , M.D.⁹, Sun Min Lee , M.D.¹⁰, Soo-Youn Lee , M.D.², and Sail Chun , M.D.¹¹; on behalf of the Clinical Mass Spectrometry Research Committee of the Korean Society of Clinical Chemistry

¹Department of Laboratory Medicine, University of Ulsan College of Medicine, Ulsan University Hospital, Ulsan, Korea; ²Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; ³Department of Laboratory Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea; ⁴Department of Laboratory Medicine, College of Medicine, The Catholic University of Korea, Seoul, Korea; ⁵Department of Laboratory Medicine, Seoul National University Hospital and College of Medicine, Seoul, Korea; ⁶Department of Laboratory Medicine and Genetics, Soonchunhyang University Bucheon Hospital, Soonchunhyang University College of Medicine, Bucheon, Korea; ⁷Department of Laboratory Medicine, Konkuk University School of Medicine, Seoul, Korea; ⁸Seoul Clinical Laboratories, Yongin, Korea; ⁹Green Cross Laboratories, Yongin, Korea; ¹⁰Department of Laboratory Medicine, Pusan National University Yangsan Hospital, Yangsan, Korea; ¹¹Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea

Background: Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is increasingly used for immunosuppressive drug tests. However, most LC-MS/MS tests are laboratory-developed and their agreement is unknown in different Korean laboratories. This interlaboratory comparison study evaluated test reproducibility and identified potential error sources.

Methods: Test samples containing three concentrations of tacrolimus, sirolimus, everolimus, cyclosporine, and mycophenolic acid were prepared by pooling surplus samples from patients undergoing routine therapeutic drug monitoring and tested in duplicate in the participating 10 clinical laboratories. Reconstitution and storage experiments were conducted for the commonly used commercial calibrator set. The robust estimators of reproducibility parameters were calculated. Spearman's rank correlation coefficient (ρ) was used to evaluate the correlation between drugs. Multiple linear regression was used to determine whether the experimental conditions alter the calibration curves.

Results: The reproducibility coefficient of variation exceeded 10% only for sirolimus concentrations 1 and 2 (10.8% and 12.5%, respectively) and everolimus concentrations 1 and 2 (12.3% and 11.4%, respectively). The percent difference values showed weak correlations between sirolimus and everolimus ($\rho=0.334$, $P=0.175$). The everolimus calibration curve slope was significantly altered after reconstitution following prolonged 5°C storage ($P=0.015$ for 14 days; $P=0.025$ for 28 days); the expected differences at 6 ng/mL were 0.598% for 14 days and 0.384% for 28 days.

Conclusions: LC-MS/MS test reproducibility for immunosuppressive drugs seems to be good in the Korean clinical laboratories. Continuous efforts are required to achieve test standardization and harmonization, especially for sirolimus and everolimus.

Key Words: Immunosuppressive drug, Therapeutic drug monitoring, Liquid chromatography coupled to tandem mass spectrometry, Interlaboratory comparison, Sirolimus, Everolimus, Reproducibility

Received: May 4, 2020

Revision received: June 12, 2020

Accepted: November 17, 2020

Corresponding author: Sail Chun, M.D.

Department of Laboratory Medicine,
University of Ulsan College of Medicine and
Asan Medical Center, 88 Olympic-ro 43-gil,
Songpa-gu, Seoul 05505, Korea
Tel: +82-2-3010-4513
Fax: +82-2-478-0884
E-mail: sailchun@amc.seoul.kr



© 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

Immunosuppressive drugs are essential for successful transplantation, as they reduce the risk of rejection [1]. The five immunosuppressive drugs most commonly prescribed for the prevention of graft loss and organ rejection are tacrolimus (TAC), sirolimus (SIR), everolimus (EVE), cyclosporine (CSA), and mycophenolic acid (MPA) [2, 3]. Calcineurin inhibitors, such as TAC, SIR, EVE, and CSA, have narrow therapeutic ranges and high interindividual pharmacokinetic variability. Therefore, therapeutic drug monitoring (TDM) is widely practiced to determine the dosage adjustments needed to reach target blood concentrations [4]. Although the utility of TDM for MPA remains controversial, monitoring MPA concentrations is also sometimes recommended [5, 6].

Reliable measurements are essential for effective TDM of immunosuppressive drugs. Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is increasingly used for TDM of immunosuppressive drugs, mainly owing to its high analytical specificity and the ability to perform multiplex tests [1]. However, most LC-MS/MS tests used in laboratories are developed, calibrated, and validated in-house. This could lead to relatively poor interlaboratory agreement between independently developed tests, thus hindering the adoption of common clinical practice guidelines [7].

There are proficiency testing (PT) programs for immunosuppressive drug TDM operated by the College of American Pathologists (CAP) or the Korean Association of External Quality Assessment Service (KEQAS). However, the CAP international PT programs do not provide nation-specific information [8], and the KEQAS PT programs included only a few participating laboratories using LC-MS/MS methods [9]. Therefore, the test reproducibility of immunosuppressive drug TDM by LC-MS/MS in Korea is unknown. An interlaboratory comparison study could determine test reproducibility across laboratories, as well as overall repeatability [10]. Interlaboratory comparisons could also contribute to laboratory performance improvement, with proper reviews of test practices and follow-up on the implementation of corrective actions [11-13].

We conducted an interlaboratory comparison study by including most of the clinical laboratories performing immunosuppressive drug TDM by LC-MS/MS in Korea and investigated calibrator reconstitution and storage conditions as potential error sources. We aimed to document current practices used for immunosuppressive drug TDM by LC-MS/MS, evaluate the reproducibility of these tests, and identify potential error sources contributing to

interlaboratory imprecision.

MATERIALS AND METHODS

Interlaboratory comparison scheme

Based on the results of a previous survey of members of the Korean Society of Clinical Chemistry [14], 10 clinical laboratories performing immunosuppressive drug TDM by LC-MS/MS were asked to participate in this study. After enrollment, the test samples were prepared in Asan Medical Center, Seoul, Korea, dispatched to the participating laboratories on dry ice in a Styrofoam box on 23 July 2018, and reached the laboratories within two days. Participating laboratories were asked to perform the test within three days of receiving the test samples. For each test sample, participating laboratories were asked to make duplicate aliquots for testing before the sample preparation step. Test results and method information, including the dates of reception and test, calibrator, internal standard, LC manufacturer, MS manufacturer and model, and extraction method, were collected in a formatted Excel file. Additional surveys regarding calibrator reconstitution protocols were conducted by e-mail. The data was compiled and analyzed at Asan Medical Center. The Institutional Review Board of Asan Medical Center exempted the approval for this study (reference number: 2018-1070).

Preparation of test samples

Surplus samples from patients undergoing routine TDM were pooled to achieve three different concentrations of test samples for each drug. The samples included EDTA whole blood for TAC, SIR, EVE, and CSA and EDTA plasma for MPA. The samples were collected for seven days and stored at 4°C up to seven days before pooling. The samples were pooled based on their reported concentrations to achieve concentrations below, within, and above the therapeutic range of each drug. Each 1 mL pooled sample was distributed in a glass vial after mixing on a roller mixer for 1 hour. The minimum number of vials for each test sample was 10; for TAC concentration 2 (TAC-2), 30 vials were prepared to evaluate homogeneity and short-term stability. The test samples were stored at -70°C before delivery. Thus, the samples underwent only one freeze-thaw cycle before being tested at the participating laboratories. For TAC-2, 10 vials were tested in triplicate on the day of production to evaluate homogeneity. To evaluate short-term stability, two vials were stored at 5°C for five days and then tested for TAC in duplicate. These four results were compared with the homogeneity values. In the homogeneity assessment, a between-sample standard deviation

$<0.3S_{\bar{x}}$ was considered acceptable. In the stability assessment, a difference from the homogeneity mean $<0.3S_{\bar{x}}$ was considered acceptable.

Details of participating laboratories

All test samples were in a frozen state on arrival at each laboratory. Details of the methods used in each laboratory are shown in Table 1. All laboratories conducted simultaneous detection with the following combinations: TAC+SIR+EVE+CSA (six laboratories) and TAC+SIR+EVE (three laboratories); one laboratory (Lab F) that was supposed to perform a CSA test with a TAC+SIR+EVE+CSA combination did not receive the CSA test samples due to a communication error. In total, TAC detection was performed in 10 laboratories, SIR detection in nine laboratories, EVE detection in nine laboratories, CSA detection in eight laboratories, and MPA detection in three laboratories. All participating laboratories used commercial calibrator sets from Chromsystems Instruments & Chemicals GmbH (Gräfelfing, Germany).

Evaluation of calibrator reconstitution method and storage as calibration bias factors

Based on the survey results of the calibrator reconstitution meth-

ods of participating laboratories, we conducted an experiment on calibrator reconstitution and storage conditions. To evaluate the impact of the calibrator reconstitution process and prolonged storage after reconstitution, we used the lyophilized calibrator 6PLUS1 Multilevel Whole Blood Calibrator Set (Lot no. 4917; Chromsystems Instruments & Chemicals GmbH) with six different concentrations plus one blank value. The calibrator contained TAC, SIR, EVE, and CSA; however, CSA could not be measured in the experimental test system.

The calibrators were stored at -70°C until use and then thawed at room temperature ($20\text{--}25^{\circ}\text{C}$) for 30 minutes before adding 2.0 mL of distilled water to each vial. The contents of the vials were then mixed under two different conditions: static incubation (sitting) for 30 minutes at room temperature ($20\text{--}25^{\circ}\text{C}$) after the addition of water followed by mixing on a roller for 60 minutes (Condition A0) or mixing on a roller for 20 minutes without static incubation (Condition B0). Aliquots from these vials were tested in four replicates. Aliquots from Condition A0 vials were initially frozen at -20°C and then thawed on different days, resulting in four storage conditions: 42 days at -20°C (Condition A1), 35 days at -20°C and seven days at 5°C (Condition A2), 28 days at -20°C and 14 days at 5°C (Condition A3), and 14 days

Table 1. Drug measurement methods used in participating laboratories

Laboratory code	Test drugs	Internal standard	LC manufacturer*	MS instrument manufacturer*	MS instrument model
A	TAC+SIR+EVE	Ascomycin+SIR- d_3 +EVE- d_4	Shimadzu	SCIEX	Triple Quad 4500
	CSA	CSD	Agilent	Agilent	6460 Triple Quadrupole
	MPA	Indomethacin	Waters	Waters	Quattro Micro
B	TAC+SIR+EVE+CSA	Ascomycin	Waters	Waters	Quattro Premier
C	TAC+SIR+EVE	Ascomycin+SIR- d_3 +EVE- d_4	Waters	Waters	Xevo TQ-S
D	TAC+CSA	Ascomycin+CSD	Waters	Waters	Xevo TQD
	MPA	Indomethacin	Waters	Waters	Xevo TQD
E	TAC+SIR+EVE	Ascomycin	Waters	Waters	Xevo TQD
	CSA	CSD	Waters	Waters	Xevo TQD
F	TAC+SIR+EVE+CSA	Ascomycin+EVE- d_4 +CSD	Agilent	Agilent	6490 Triple Quadrupole
G	TAC+SIR+EVE+CSA	Ascomycin+SIR- d_3 +EVE- d_4 +CSD	Agilent	SCIEX	API 4000
H	TAC+SIR+EVE+CSA	TAC- ^{13}C , d_2 +SIR- d_3 +EVE- d_4 +CSA- d_4	Agilent	SCIEX	API 3200
I	TAC+SIR+EVE+CSA	Ascomycin	Agilent	SCIEX	QTRAP 5500
	MPA	MPA- d_3	Agilent	SCIEX	QTRAP 5500
J	TAC+SIR+EVE+CSA	Ascomycin+CSD	Agilent	SCIEX	Triple Quad 3500

Most laboratories used precipitation with organic solvent mixture followed by centrifugation for extraction, except Lab I, which used liquid/liquid extraction. All the participating laboratories used the commercial calibrator from Chromsystems Instruments & Chemicals GmbH (Gräfelfing, Germany).

*The manufacturer names are indicated using their short names: Agilent, Agilent Technologies (Santa Clara, CA, USA); SCIEX, AB SCIEX (Foster City, CA, USA); Shimadzu, Shimadzu Corporation (Kyoto, Japan); Waters, Waters Corporation (Milford, MA, USA).

Abbreviations: CSA, cyclosporine; CSD, cyclosporine D; EVE, everolimus; LC, liquid chromatography; MPA, mycophenolic acid; MS, mass spectrometry; TAC, tacrolimus; SIR, sirolimus.

at -20°C and 28 days at 5°C (Condition A4). These aliquots were tested in duplicate on the same day.

Simultaneous quantification of TAC, SIR, and EVE was performed for the calibrators in the same manner as for patient samples at Asan Medical Center. Briefly, 40 µL samples were prepared by protein precipitation using 80 µL of aqueous 0.1 M zinc sulfate and 200 µL of acetonitrile containing the internal standards. The internal standard solution consisted of ascomycin (Cerilliant, Round Rock, TX, USA), SIR-d₃ (IsoSciences, Ambler, PA, USA), and EVE-d₄ (Cerilliant). HPLC was conducted using an ACQUITY UPLC I-Class system (Waters Corporation, Milford, MA, USA) equipped with an ACQUITY UPLC HSS SB C18 column (30 mm × 2.1 mm, 4 µm; Waters Corporation). The system was operated at a flow rate of 0.2 mL/min with a column temperature of 55°C. The mobile phase for isocratic elution consisted of 50% methanol (HPLC grade; Merck, Darmstadt, Germany) and 50% ammonium acetate buffer (2 mM, pH 2.7, with ≥98% acetate ammonium; Sigma-Aldrich, St. Louis, MO, USA). MS was detected on a Xevo TQ-S LC-MS/MS system (Waters Corporation) using positive electrospray ionization and argon as the collision gas. Peak areas were obtained by multiple reaction monitoring of the following mass transitions for quantification: TAC (m/z 821.5 > 768.5), SIR (m/z 931.6 > 864.6), EVE (m/z 975.7 > 908.6), ascomycin (m/z 809.5 > 756.5), SIR-d₃ (m/z 934.6 > 864.6), and EVE-d₄ (m/z 979.7 > 912.6). Response values were calculated using the resultant peak areas (area of analyte/area of internal standard). LC-MS/MS analysis and data acquisition were managed using Masslynx software version 4.1 (Waters Corporation). TargetLynx software version 4.1 (Waters Corporation) was used for chromatographic peak detection and baseline determination.

Statistical analysis

Data from the participating laboratories were used to estimate test reproducibility. The repeatability standard deviation (S_r) is the root mean square value of all differences in w_i , where w_i is the difference between the duplicate measurement results in the i^{th} laboratory and the number of participating laboratories is p :

$$S_r = \sqrt{\frac{\sum_{i=1}^p w_i^2}{p}}$$

The standard deviation of the sample averages ($S_{\bar{x}}$) is:

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^p d_i^2}{(p-1)}}$$

where $d_i = \bar{x}_i - \bar{\bar{x}}$ and the average of the duplicate measurement results in the i^{th} laboratory is \bar{x}_i , and the average of the \bar{x} for one test sample is $\bar{\bar{x}}$.

When $(S_{\bar{x}})^2 > (S_r)^2/2$, the reproducibility standard deviation (S_R) is:

$$S_R = \sqrt{(S_{\bar{x}})^2 + (S_r)^2/2}$$

The reproducibility coefficient of variation (CV_R) is:

$$CV_R = 100S_R/\bar{\bar{x}}$$

Robust statistical methods based on the medians were also used to describe the result distribution. The robust repeatability standard deviation (S_r^*) was calculated using Algorithm S (ISO-13528:2015) [15, 16]. The robust standard deviation of the sample averages ($S_{\bar{x}}^*$) is:

$$S_{\bar{x}}^* = 1.483\text{med}(|D_i|)$$

which is the scaled median absolute deviation ($MADe$), where $D_i = \bar{x}_i - \text{med}(\bar{x})$. The percent difference ($D\%$) was used to describe the deviation in the individual laboratory results, calculated as $\%D_i = 100D_i/\text{med}(\bar{x})$. The robust reproducibility standard deviation (S_R^*) and the robust reproducibility CV (CV_R^*) were calculated as $S_R^* = \sqrt{(S_{\bar{x}}^*)^2 + (S_r^*)^2/2}$ and $\%RSD_R^* = 100S_R^*/\text{med}(\bar{x})$, respectively. Spearman's rank correlation coefficient (ρ) was calculated to determine the relationship between the $D\%$ values of different test drugs paired with similar drug concentration measurements.

For the result analysis, a multiple linear regression model was used to compare the slopes of the calibration curves (non-weighted):

$$Y = \beta_0 + \beta_1 X + \beta_2 X D_c + \varepsilon$$

where the response value is Y , the assigned value of calibrator is X , and the dummy variable of the conditions is D_c . The t -test on coefficient β_2 was performed to determine whether the slopes obtained from the reconstitution or the storage conditions were different from those at the baseline conditions (A0 or A1), with $P \leq 0.05$. Data integration, analysis, and visualization were performed using R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Homogeneity and short-term stability of TAC-2

In the homogeneity test, the overall average of TAC-2 was 8.99 ng/mL. Because the variance of sample averages was $< 1/3$ of the within-sample variance, the between-sample standard deviation was estimated as 0. Therefore, the homogeneity was considered sufficient. In the short-term stability test, the overall average of TAC-2 was 8.88 ng/mL. Its difference from the homogeneity mean was 0.11 ng/mL, which was lower than the calcu-

lated check value (0.17, 0.3S_{x̄} of TAC-2). Therefore, the stability was considered acceptable.

Measurement results of the participating laboratories

Statistical summary of the laboratory measurement results is

shown in Table 2. The robust reproducibility CVs (CV_R^{*}) were >10% for SIR-1, SIR-2, EVE-1, and EVE-2. A pattern suggesting a positive correlation among test samples was identified when the distribution of D% values was visualized using a heatmap table (Fig. 1A). For correlation analysis between the results for different

Table 2. Summary statistics of laboratory measurement results

Drug	Participating laboratories (N)	Test sample	General without outlier removal					Robust statistics				
			Mean (\bar{x})	S _r	S _{z̄}	S _R	CV _R , %	Median (med(\bar{x}))	S _r [*]	S _{z̄} [*]	S _R [*]	CV _R [*] , %
TAC (ng/mL)	10	TAC-1	5.41	0.41	0.48	0.57	10.5	5.59	0.31	0.24	0.32	5.8
		TAC-2	8.93	0.39	0.57	0.63	7.1	8.84	0.24	0.19	0.25	2.8
		TAC-3	15.83	0.50	1.33	1.37	8.7	16.05	0.30	1.43	1.45	9.0
SIR (ng/mL)	9	SIR-1	3.50	0.19	0.29	0.32	9.1	3.55	0.14	0.37	0.38	10.8
		SIR-2	5.83	0.36	0.88	0.92	15.7	5.55	0.27	0.67	0.69	12.5
		SIR-3	12.08	1.13	0.61	1.00	8.3	11.96	0.85	0.61	0.86	7.2
EVE (ng/mL)	9	EVE-1	2.43	0.25	0.33	0.38	15.5	2.45	0.19	0.27	0.30	12.3
		EVE-2	3.53	0.39	0.36	0.46	12.9	3.55	0.27	0.36	0.40	11.4
		EVE-3	5.89	0.25	0.49	0.53	8.9	5.90	0.19	0.52	0.54	9.1
CSA (ng/mL)	8	CSA-1	30.10	1.67	2.83	3.06	10.2	29.85	0.78	0.70	0.89	3.0
		CSA-2	63.78	2.32	3.77	4.12	6.5	63.45	1.15	2.37	2.51	4.0
		CSA-3	149.93	4.55	10.44	10.92	7.3	149.40	3.53	9.01	9.35	6.3
MPA (mg/L)	3	MPA-1	0.76	0.06	0.06	0.07	9.3	0.77	0.04	0.07	0.07	9.6
		MPA-2	1.94	0.14	0.18	0.21	10.7	1.90	0.11	0.17	0.19	9.8
		MPA-3	5.16	0.08	0.12	0.14	2.6	5.10	0.06	0.04	0.06	1.1

Abbreviations: CSA, cyclosporine; EVE, everolimus; MPA, mycophenolic acid; TAC, tacrolimus; SIR, sirolimus.

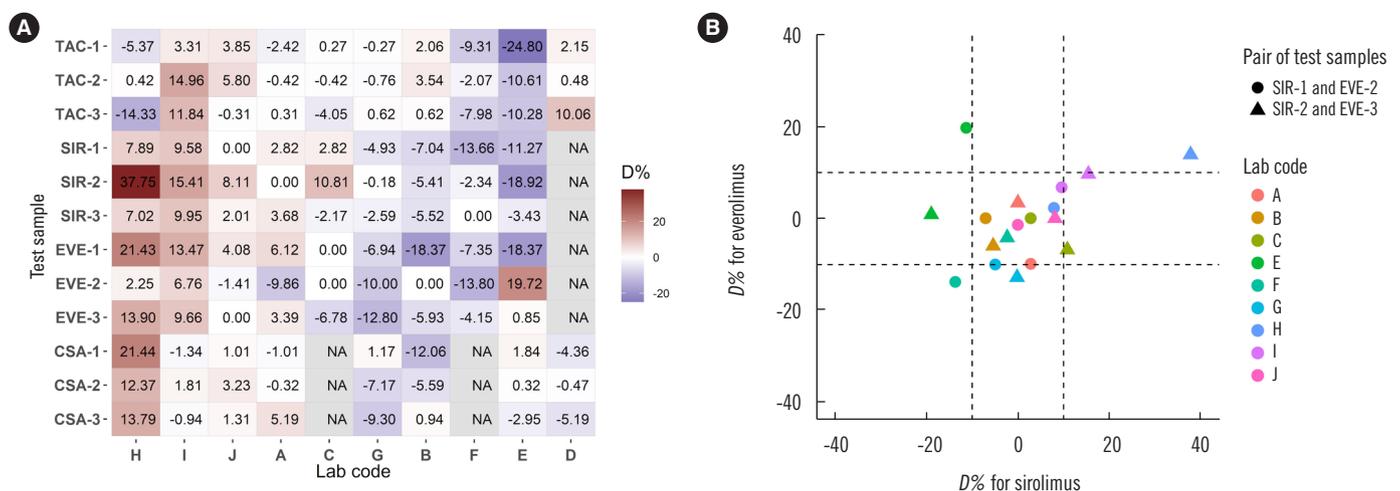


Fig. 1. Correlation between the test samples. (A) The percent difference in individual laboratory results (D%) for each test sample. The cells are colored according to the D% values. The laboratory order is based on the average of the D% values, except for Lab E and Lab D, for pattern readability. (B) Correlation of D% values for test samples with similar concentrations for SIR and EVE. Test sample pairs are indicated by a triangle or circle, and the colors indicate the laboratory code.

Abbreviations: CSA, cyclosporine; EVE, everolimus; NA, not available; SIR, sirolimus; TAC, tacrolimus.

drugs, the test samples were paired in similar concentrations: SIR-1 and EVE-2 and SIR-2 and EVE-3 (N=18); TAC-1 and SIR-2 (N=9); and TAC-1 and EVE-3 (N=9). The ρ values between SIR and EVE, TAC and SIR, and TAC and EVE results were 0.334 ($P=0.175$), 0.383 ($P=0.308$), and -0.150 ($P=0.700$), respectively. SIR and EVE results showed a weak positive correlation (Fig. 1B).

Table 3. Additional calibrator reconstitution survey results

Laboratory code	Static incubation time, min	Mix time, min	Mixing by	Use of timer (Static/Mix)
A	30	60	Roller mixer	Yes/Yes
B	15	15	Roller mixer	Yes/Yes
C	15	5	Roller mixer	Yes/Yes
D	10	20	Roller mixer	Yes/Yes
E	0	60	Roller mixer	Yes
F	20–30	10	Roller mixer	No/Yes
G	0	45	Roller mixer	Yes
H	0	20	Roller mixer	Yes
I	10–15	20–30	Roller mixer	No/No
J	0	90	Shaker	Yes

Calibrator reconstitution and storage conditions experiment

Additional survey results for the calibrator reconstitution protocols are shown in Table 3. Based on the survey results, Conditions A0 (similar to Lab A protocol) and B0 (similar to Lab H protocol) were selected. The slopes of the calibration curves did not significantly differ between these two conditions (Table 4).

In the experiment on storage conditions, the calibration curve slopes of SIR from Conditions A2 and A3 ($P<0.001$ and $P=0.030$) and of EVE from Conditions A3 and A4 ($P<0.001$ and $P=0.009$) were significantly different from those of Condition A1. When the response value corresponding to a concentration, for example, of 6 ng/mL (similar to the mean concentrations of SIR-2) from the baseline condition calibration curve was entered into the experimental condition equation, positive biases were present in conditions A1–A3; however, the percent differences were expected to be in the range of 0.217–0.598% (Table 4).

DISCUSSION

This is the first interlaboratory comparison study of LC-MS/MS for immunosuppressive drug TDM in Korea. The test reproducibility was generally good, and the robust reproducibility CVs were

Table 4. Calibrator reconstitution and storage experiment results

Experiment	Drug	Condition	Coefficient β_2 (95% CI)	P^*	Expected concentration at 6 ng/mL (%diff)
Reconstitution	TAC	A0	Baseline		6.000 (Baseline)
		B0	0.00012 (-0.00034–0.00058)	0.612	5.997 (-0.049)
	SIR	A0	Baseline		6.000 (Baseline)
		B0	-0.00005 (-0.00043–0.00032)	0.772	6.002 (0.029)
	EVE	A0	Baseline		6.000 (Baseline)
		B0	-0.00015 (-0.00036–0.00005)	0.141	6.007 (0.112)
Storage	TAC	A1	Baseline		6.000 (Baseline)
		A2	-0.00037 (-0.00102–0.00028)	0.257	6.009 (0.143)
		A3	-0.00028 (-0.00093–0.00037)	0.386	6.007 (0.109)
		A4	-0.00022 (-0.00087–0.00043)	0.496	6.005 (0.086)
	SIR	A1	Baseline		6.000 (Baseline)
		A2	-0.00157 (-0.00239–0.00076)	<0.001	6.033 (0.555)
		A3	-0.00091 (-0.00173–0.00009)	0.030	6.019 (0.321)
		A4	-0.00061 (-0.00143–0.00020)	0.138	6.013 (0.217)
	EVE	A1	Baseline		6.000 (Baseline)
		A2	-0.00043 (-0.00098–0.00012)	0.121	6.013 (0.222)
		A3	-0.00117 (-0.00172–0.00062)	<0.001	6.036 (0.598)
		A4	-0.00075 (-0.00130–0.00020)	0.009	6.023 (0.384)

* P from t -test on coefficient β_2 .

Abbreviations: CI, confidence interval; EVE, everolimus; TAC, tacrolimus; SIR, sirolimus.

slightly greater than 10% for SIR-1, SIR-2, EVE-1, and EVE-2 with low analyte concentrations. The storage conditions experiment showed that the calibration curve slopes for SIR and EVE changed with prolonged storage of the reconstituted calibrator at 5°C; however, the magnitude of the change was small. Thus, other factors might affect their interlaboratory imprecision.

The method survey results showed that the method details varied across laboratories. For sample preparation, most laboratories used simple protein precipitation methods, and none used additional solid-phase extraction. For internal standards, structural analogs, such as ascomycin, cyclosporine D, and indomethacin, were more commonly used than a stable isotope-labeled internal standard. Although some laboratories used similar sample preparation methods and internal standards and the same commercial calibrator, their test performance could vary depending on the specific protocol followed [17].

In our study, the robust reproducibility CVs were >10% for the SIR-1, SIR-2, EVE-1, and EVE-2 test samples. Similarly, in the participant summary reports of the CAP international PT programs conducted in 2018, the CVs of SIR and EVE measured by LC-MS/MS were higher than those of the other drugs [8, 18-20]. The reconstitution procedure of the commercial calibrator from Chromsystems is not specified by the manufacturer in detail [21], and the reported times for static incubation and mixing varied across laboratories (Table 3). Considering this, we experimentally determined whether calibrator handling affects the calibration curve and contributes to interlaboratory imprecision. In our experiment, reconstitution with a shorter mixing time did not significantly alter the calibration curve slope, suggesting that the impact of the differences in mixing times is relatively small. However, to prevent possible inadequate mixing or unnecessary time-consuming procedures, the manufacturer's recommendation for mixing the calibrator should be specified more clearly. Regarding storage conditions, the manufacturer recommends that the reconstituted calibrator can be stored up to one week at 2–8°C. In our experiment, the calibration curve slopes for SIR and EVE changed with prolonged storage (>one week) of the reconstituted calibrator at 5°C. However, the magnitude of the difference was small, suggesting that a certain extent of prolonged storage after reconstitution might not be a major source of bias.

Although the use of a common calibrator for LC-MS/MS measurement of serum 25-hydroxyvitamin D significantly improved interlaboratory agreement [22], sufficient harmonization among LC-MS/MS methods cannot be achieved by simply using the same commercial calibrator set, as demonstrated for TAC by Levine, *et al.* [23]. The use of a stable isotope-labeled internal

standard, which is generally considered the most effective way of compensating for the matrix effect, also does not ensure test performance improvement [24, 25]. This is because LC-MS/MS tests can be affected by other factors, such as sample and reagent volumes, sample preparation protocols, chromatography procedure, and the mass transitions monitored [7]. The interlaboratory comparison in the ongoing Zortracker EVE study is a good effort in this regard; improved interlaboratory CV has been demonstrated across the LC-MS/MS laboratories during the observation period, along with a decreased number of outliers [26].

Our study has several limitations: first, we did not assign a reference value to the test samples; therefore, we could not assess laboratory bias. Instead, we aimed to assess test reproducibility in this first interlaboratory comparison in Korea. A refined follow-up study with traceable assigned values of test samples is needed to assess laboratory bias. Our results could serve as the basis for such a study. In addition, we could not evaluate the homogeneity and the stability of all the test samples owing to the insufficient sample volume. However, we produced the test samples at the same time and in the same manner, and TAC-2 representatively showed acceptable homogeneity. To minimize potential issues with stability, the test samples were shipped in a frozen state and were thawed on the day that the laboratories conducted the tests.

It is challenging to reliably assess interlaboratory comparison results with a small number of participating laboratories [27]. Estimation of location and dispersion for very small datasets can be biased to some degree, even with reasonable outlier rejection and robust statistics. The possibility of high variability in dispersion estimates should be noted. For a balanced summary, we described both estimators using general statistics without outlier removal or robust statistics. As most people in charge of LC-MS/MS in clinical laboratories are members of the Korean Society of Clinical Chemistry, all of whom participated in this study, the possibility of sampling bias is minimal.

This study provides a snapshot of the current status of immunosuppressive drug TDM by LC-MS/MS in Korea. Although the test reproducibility seems to be good, continuous efforts are required to achieve test standardization and harmonization, especially for SIR and EVE. This study could provide background data for further interlaboratory comparison studies and PT. Unlike general chemistry tests, immunosuppressive drug TDM by LC-MS/MS is conducted as a multiplex test by a small number of laboratories. Further studies should be designed with consideration of these conditions.

ACKNOWLEDGEMENTS

The authors thank the staff in the participating laboratories for their cooperation and the sample testing.

AUTHOR CONTRIBUTIONS

Kim HK collected and analyzed the data and drafted the manuscript. Lee SG, Chae H, Song SH, Ahn S, Kim S, and Lee SM analyzed the data and interpreted the results. Park HD, Lee YH, Yun YM, and Lee SY contributed to the conception. SC designed and supervised the study. All authors acquired data from each institution, reviewed the manuscript, and provided critical feedback.

CONFLICTS OF INTEREST

No potential conflicts of interest relevant to this article are reported.

RESEARCH FUNDING

This study was supported by the Clinical Mass Spectrometry Research Committee of the Korean Society of Clinical Chemistry.

ORCID

Hyun-Ki Kim	https://orcid.org/0000-0002-3299-5298
Hyung-Doo Park	https://orcid.org/0000-0003-1798-773X
Sang-Guk Lee	https://orcid.org/0000-0003-3862-3660
Hyojin Chae	https://orcid.org/0000-0001-5870-1887
Sang Hoon Song	https://orcid.org/0000-0002-5084-1137
Yong-Wha Lee	https://orcid.org/0000-0002-4459-6286
Yeo-Min Yun	https://orcid.org/0000-0002-5485-8331
Sunhyun Ahn	https://orcid.org/0000-0003-4741-6903
Serim Kim	https://orcid.org/0000-0003-0269-5099
Sun Min Lee	https://orcid.org/0000-0002-2896-3365
Soo-Youn Lee	https://orcid.org/0000-0001-7595-4042
Sail Chun	https://orcid.org/0000-0002-5792-973X

REFERENCES

- McShane AJ, Bunch DR, Wang S. Therapeutic drug monitoring of immunosuppressants by liquid chromatography-mass spectrometry. *Clin Chim Acta* 2016;454:1-5.
- Zhang Y and Zhang R. Recent advances in analytical methods for the therapeutic drug monitoring of immunosuppressive drugs. *Drug Test Anal* 2018;10:81-94.
- Khan S, Khan S, Baboota S, Ali J. Immunosuppressive drug therapy-biopharmaceutical challenges and remedies. *Expert Opin Drug Deliv* 2015;12:1333-49.
- Mohammadpour N, Elyasi S, Vahdati N, Mohammadpour AH, Shamsara J. A review on therapeutic drug monitoring of immunosuppressant drugs. *Iran J Basic Med Sci* 2011;14:485-98.
- Bentata Y. Mycophenolates: the latest modern and potent immunosuppressive drugs in adult kidney transplantation: what we should know about them? *Artif Organs* 2020;44:561-76.
- Kuypers DR, Le Meur Y, Cantarovich M, Tredger MJ, Cattaneo D, et al. Consensus report on therapeutic drug monitoring of mycophenolic acid in solid organ transplantation. *Clin J Am Soc Nephrol* 2010;5:341-58.
- Annesley TM, McKeown DA, Holt DW, Mussell C, Champarnaud E, Harter L, et al. Standardization of LC-MS for therapeutic drug monitoring of tacrolimus. *Clin Chem* 2013;59:1630-7.
- CAP. Immunosuppressive drugs. CSM-A 2018; participant summary. Northfield, IL: College of American Pathologists. 2018.
- KEQAS. Immunosuppressants therapeutic drug monitoring 2018; summary report. Korean Association of External Quality Assessment Service. 2018.
- Berlinger B and Harper M. Interlaboratory comparison for the determination of the soluble fraction of metals in welding fume samples. *J Occup Environ Hyg* 2018;15:152-6.
- Christians U, Vinks AA, Langman LJ, Clarke W, Wallemacq P, van Gelder T, et al. Impact of laboratory practices on interlaboratory variability in therapeutic drug monitoring of immunosuppressive drugs. *Ther Drug Monit* 2015;37:718-24.
- Pitkälä A, Gindonis V, Wallin H, Honkanen-Buzalski T. Interlaboratory proficiency testing as a tool for improving performance in laboratories diagnosing bovine mastitis. *J Dairy Sci* 2005;88:553-9.
- Jeong TD, Lee H, 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.
- Chae H, Cho SE, Park HD, Chun S, Lee YW, Yun YM, et al. Use of liquid chromatography-tandem mass spectrometry for clinical testing in Korean laboratories: a questionnaire survey. *Ann Lab Med* 2019;39:447-53.
- Volodarsky ET, Warsza ZL, Kosheva L, Idzkowski A. Robust algorithm S to assess the precision of interlaboratory measurements. *Meas Autom Monit* 2015;61:111-4.
- ISO. Statistical methods for use in proficiency testing by interlaboratory comparison. ISO 13528:2015. Geneva: International Organization for Standardization. 2015.
- Honour JW. Development and validation of a quantitative assay based on tandem mass spectrometry. *Ann Clin Biochem* 2011;48:97-111.
- CAP. Everolimus. EV-A 2018; participant summary. Northfield, IL: College of American Pathologists. 2018.
- CAP. Everolimus. EV-B 2018; participant summary. Northfield, IL: College of American Pathologists. 2018.
- CAP. Immunosuppressive drugs. CSM-B 2018; participant summary. Northfield, IL: College of American Pathologists. 2018.
- Chromsystems. 6PLUS1 multilevel calibrator set, immunosuppressants. 28039. Gräfelfing: Chromsystems Instruments and Chemicals GmbH. 2018.
- Carter GD and Jones JC. Use of a common standard improves the performance of liquid chromatography-tandem mass spectrometry methods for serum 25-hydroxyvitamin-D. *Ann Clin Biochem* 2009;46:79-81.
- Levine DM, Maine GT, Armbruster DA, Mussell C, Buchholz C, O'Connor G, et al. The need for standardization of tacrolimus assays. *Clin Chem*

- 2011;57:1739-47.
24. Valbuena H, Shipkova M, Kliesch SM, Müller S, Wieland E. Comparing the effect of isotopically labeled or structural analog internal standards on the performance of a LC-MS/MS method to determine ciclosporin A, everolimus, sirolimus and tacrolimus in whole blood. *Clin Chem Lab Med* 2016;54:437-46.
 25. Zhao Y, Liu G, Shen JX, Aubry AF. Reasons for calibration standard curve slope variation in LC-MS assays and how to address it. *Bioanalysis* 2014; 6:1439-43.
 26. Schniedewind B, Meyer EJ, Christians U. Long-term performance of laboratory-developed liquid chromatography-tandem mass spectrometry tests and a Food and Drug Administration-approved immunoassay for the therapeutic drug monitoring of everolimus. *Ther Drug Monit* 2020; 42:421-6.
 27. Wong SK. Performance evaluation for proficiency testing with a limited number of participants. *Accred Qual Assur* 2011;16:539.