Research Article

Retrospective Data Analysis and Proposal of a Practical Acceptance Criterion for Inter-laboratory Cross-validation of Bioanalytical Methods Using Liquid Chromatography/Tandem Mass Spectrometry

Tomoki Yoneyama,^{1,3} Takashi Kudo,¹ Fumihiko Jinno,¹ Eric R. Schmidt,² and Takahiro Kondo¹

Received 12 May 2014; accepted 29 July 2014; published online 15 August 2014

Abstract. The purpose of this study is to conduct a retrospective data analysis for inter-laboratory cross-validation studies to set a reasonable and practical acceptance criterion based on a number of cross-validation results. From the results of cross-validation studies for 16 compounds and their metabolites, analytical bias and variation were evaluated. The accuracy of cross-validation samples was compared with that of quality control (QC) samples with statistical comparison of the analytical variation. An acceptance criterion was derived with a confidential interval approach. As the results, while a larger bias was observed for the cross-validation samples, the bias was not fully caused by analytical variation or bias attributable to the analytical methods. The direction of the deviation between the cross-validation samples and QC samples was random and not concentration-dependent, suggesting that inter-laboratory variability such as preparation errors could be a source of bias. A derived acceptance criterion corresponds to one prescribed in the Guideline on bioanalytical method validation from the Ministry of Health, Labour and Welfare in Japan and is a little wider than one in the European Medical Agency. In conclusion, thorough retrospective data analysis revealed potential causes of larger analytical bias in inter-laboratory cross-validation studies. A derived acceptance criterion would be practical and reasonable for the inter-laboratory cross-validation study.

KEY WORDS: acceptance criterion; cross-validation; Guideline on bioanalytical method validation (BMV); inter-laboratory reliability; regulated bioanalysis.

INTRODUCTION

Quantitative determination of drugs and their metabolites in biological matrices is an integral part of drug research and development for evaluation of the safety and efficacy of new pharmaceutical substances through the elucidation of the pharmacokinetics (absorption, distribution, metabolism, and excretion) of drugs and their metabolites. The concentration data are also used in the evaluation and interpretation of bioavailability, bioequivalence, and drug–drug interaction findings. It is therefore essential to employ bioanalytical methods that have been shown to be robust and well characterized through full validation of the bioanalytical methods in order to generate reliable and reproducible results that can be satisfactorily interpreted.

Under these circumstances, the principles for the validation of bioanalytical methods have been well discussed at the American Association of Pharmaceutical Scientists (AAPS)/

Food and Drug Administration (FDA) Bioanalytical Workshops held in 1990 (1), 2000 (2), 2007 (3), and 2013. These workshops not only provided a platform for scientific discussions on this theme, but also formed the basis for the regulatory guidance on bioanalytical method validation (BMV) issued by the FDA in 2001 (4) and the draft FDA guidance on BMV in 2013 (5). During this period, the European Medical Agency (EMA) also released their own Guideline on BMV in 2011 (6). In Japan, a Draft guideline on BMV in Japan by Japan Bioanalysis Forum was published in 2012 (7), followed by the Guideline on BMV in pharmaceutical development which was released from the Japanese authority, Ministry of Health, Labour and Welfare (MHLW) in 2013 (8).

Cross-validation is one of key principles of BMV where the validation parameters and the results obtained from two or more bioanalytical methods are compared. When sample analyses within a single study are conducted at more than one site or more than one laboratory, cross-validation should be conducted to establish inter-laboratory reliability. Although the need of cross-validation study had been well discussed from the 1st AAPS/FDA Bioanalytical Workshop (1), no specific procedure and criterion for cross-validation was prescribed in the FDA Guidance in 2001 (4). Followed by the 3rd AAPS/FDA Bioanalytical Workshop (3) where the attendees agreed that cross-validation procedure and acceptance criteria need to remain flexible considering the various

¹ Drug Metabolism and Pharmacokinetics Research Laboratories, Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, 26-1, Muraoka-higashi 2-chome, Fujisawa, Kanagawa 251-8555, Japan.

² Clinical Pharmacology, Clinical Science, Takeda Development Center Americas, Inc., Deerfield, Illinois, USA.

³ To whom correspondence should be addressed. (e-mail: tomoki.yoneyama@takeda.com)

bioanalytical situations where they would be required, no detailed criterion for cross-validation study was provided in the FDA draft guidance in 2013 (5).

The EMA Guideline on BMV (6) is the first regulatory guideline which provides specific criteria for cross-validation using the quality control (QC) samples, where the obtained mean accuracy by the different methods should be within 15% and may be wider, if justified for the QC samples. While the provided cutoff value of 15% is well aligned with the criteria for accuracy evaluation of QC samples during the full validation of bioanalytical methods, it seems that the guideline also leaves room for employing a wider acceptance criterion considering the various situations of cross-validation experiments. In case of the inter-laboratory cross-validation study, especially, the use of wider acceptance criterion may be reasonable since additional variation could be introduced by the involvement of at least one other bioanalytical laboratory. The recently released MHLW Guideline on BMV (8) also provides an acceptance criterion for inter-laboratory cross-validation study: The mean accuracy of QC sample should be within $\pm 20\%$ deviation of the theoretical concentration, considering intra- and inter-laboratories precision. Although several articles have discussed the general approach and acceptance criterion on cross-validation (9–13), they have not been based on a large number of actual cross-validation results.

The purpose of this study is to conduct a retrospective data analysis of the inter-laboratory cross-validation studies for 16 compounds and their metabolites to set a reasonable and practical acceptance criterion. A further investigation is performed for the larger analytical bias observed for cross-validation samples during this data analysis. In addition, a practical and reasonable acceptance criterion for inter-laboratory cross-validation studies is derived based on a number of actual inter-laboratory cross-validation results.

MATERIALS AND METHODS

Compound Property and Bioanalytical Methods

The 16 compounds evaluated in this data analysis were candidate compounds under clinical development in multiple countries by Takeda Pharmaceutical Company Limited. The inter-laboratory cross-validation studies for the 16 compounds and their metabolites were performed from 2005 to 2012. The physicochemical properties of the compounds and their metabolites are summarized in Table I. Bioanalytical methods used for the cross-validation studies are summarized in Table II. The bioanalytical methods for the 16 compounds and their metabolites were developed for human plasma and urine as target biological matrices of interest using liquid chromatography/tandem mass spectrometry (LC/MS/MS). All the bioanalytical methods were fully validated at both the original and receiving laboratories, individually, in accordance with the FDA guidance on BMV. Original laboratory means the bioanalytical laboratory where the cross-validation samples were prepared and subsequently shipped to a receiving laboratory. Both original and receiving laboratories were bioanalytical CROs in our inter-laboratory cross-validation studies. In most cases, a bioanalytical method was developed at the in-house laboratory and then transferred and validated at a bioanalytical laboratory. After completion of validation at the original laboratory, the method

was transferred to a second bioanalytical laboratory, the receiving laboratory. The validated bioanalytical methods were applied to the sample analyses in the cross-validation studies. While the bioanalytical methods were not identical between original and receiving laboratories, they were basically very similar as the bioanalytical methods validated at one laboratory were then transferred to the other. No obvious bioanalytical issue was observed at either the original or the receiving laboratories during the full validation of the bioanalytical methods used for the cross-validation studies.

Procedure of the Inter-laboratory Cross-validation Studies

Two sets of spiked cross-validation samples at three concentration levels (low, medium, and high concentration levels) were prepared at the original laboratory. One set of spiked cross-validation samples was analyzed in triplicate at the original laboratory. Another set of cross-validation samples was shipped to the receiving laboratory with temperature monitored and then analyzed in triplicate within established stability period. The concentrations of the cross-validation samples were kept blinded to the receiving laboratory up to the completion of the cross-validation studies. The cross-validation samples were analyzed by fully validated bioanalytical methods at both original and receiving laboratories along with calibration standards and QC samples for the validity of each analytical run. The QC samples at three concentration levels were analyzed in duplicate. The acceptance of each analytical run was evaluated based on the accuracy of the calibration standards and QC samples as per the FDA guidance on BMV (4). The acceptance criterion taken for our cross-validation studies was that the accuracy of two or more cross-validation samples at each concentration level should be within $\pm 20\%$ of their respective nominal concentration at both the original and receiving laboratories.

Data Analysis

Mean accuracy, absolute value of mean relative error ($|\text{mean RE}|$), standard deviation (SD), and coefficient of variation (CV) were calculated at each concentration level for the cross-validation samples and QC samples. Mean, SD, and CV of the mean accuracy as well as the mean of $|\text{mean RE}|$ and CV were also calculated. All the calculations and *F*-test were performed using Microsoft Excel 2007.

Comparison of the Accuracy Between the Cross-validation Samples and QC Samples

The accuracy between cross-validation samples and QC samples was compared by statistically comparing the analytical variation. The range of the analytical variation that could statistically include 99% of the individual accuracy data (the range between 0.5 and 99.5th percentile), “the 99% range of accuracy,” hereafter, was calculated with the Eq. (1).

$$\text{The 99\% range of accuracy (\%)} = \text{mean accuracy} \pm 2.58 \times \text{SD} \quad (1)$$

The 99% range of the accuracy was calculated for all of analytes at three concentration levels analyzed at both the

Table I. Physicochemical Properties of Compounds and Their Metabolites

Name	Molecular weight	Acid/base/neutral	cLogP	Topological polar surface area	Hydrogen bond donor/acceptor
Compound A	339	Base	1.0	97	1/5
Metabolite A-1	325	Base	0.4	108	2/5
Compound B	480	Neutral	1.3	107	2/8
Metabolite B-1	496	Neutral	0.8	127	3/9
Metabolite B-2	494	Neutral	-0.5	124	2/9
Metabolite B-3	478	Neutral	1.7	104	1/8
Metabolite B-4	315	Acid	1.4	92	2/5
Compound C	624	Neutral	4.6	133	2/10
Compound D	569	Acid	4.2	156	1/7
Metabolite D-1	457	Acid	4.9	123	2/6
Metabolite D-2	415	Zwitterionic	4.8	114	3/6
Metabolite D-3	428	Acid	3.9	134	3/5
Compound E	500	Base	3.2	89	1/8
Metabolite E-1	486	Base	2.5	100	2/8
Compound F	374	Base	0.3	106	2/5
Metabolite F-1	390	Base	-0.2	125	2/6
Compound G	345	Base	2.5	64	1/4
Metabolite G-1	346	Acid	3.2	89	1/5
Metabolite G-2	205	Acid	3.0	53	2/2
Metabolite G-3	359	Acid	2.4	78	0/3
Metabolite G-4	442	Neutral	1.2	130	2/8
Compound H	259	Neutral	2.5	38	1/3
Metabolite H-1	291	Acid	1.0	87	3/5
Metabolite H-2	275	Neutral	2.1	59	2/4
Metabolite H-3	273	Neutral	1.2	55	1/4
Metabolite H-4	289	Neutral	0.8	76	2/5
Compound I	424	Neutral	5.1	59	2/4
Compound J	222	Base	1.2	50	1/5
Metabolite J-1	208	Base	1.1	59	2/5
Metabolite J-2	166	Base	-0.5	73	2/5
Metabolite J-3	254	Zwitterionic	-1.7	102	2/7
Compound K	381	Neutral	6.3	25	0/4
Compound L	1,225	Base	-2.5	476	18/29
Compound M	504	Base	1.7	109	3/6
Compound N	525	Acid	4.7	99	1/7
Metabolite N-1	362	Acid	3.2	81	1/5
Compound O	298	Base	4.9	15	1/3
Metabolite O-1	328	Zwitterionic	2.0	53	2/5
Metabolite O-2	314	Base	4.3	36	2/4
Compound P	283	Base	4.5	12	1/2
Metabolite P-1	313	Zwitterionic	1.5	49	2/4

original and receiving laboratories. The number of QC samples whose accuracy was out of the 99% range of accuracy was then counted.

For visualization purpose, the mean accuracy of cross-validation samples, individual accuracy of QC samples, and the 99% range of accuracy were normalized according to the following equations.

The normalized accuracy (%)

$$= \frac{\text{accuracy}}{\text{mean accuracy of cross validation samples}} \times 100 \quad (2)$$

The normalized 99% range of accuracy (%) = $100 \pm 2.58 \times CV$

(3)

RESULTS

Characteristics of the Compounds and Bioanalytical Methods

The physicochemical properties of the 16 compounds and their metabolites are summarized in Table I. The compound set included 15 bases, 10 acidic compounds, 12 neutral compounds, and 4 zwitterionic compounds. Molecular weights ranged from 166 to 1,225 Da, the cLogP ranged from -2.5 to 6.3, and the topographical polar surface area ranged from 12 to 476.

The summary of bioanalytical methods used for inter-laboratory cross-validation studies is presented in Table II. Five bioanalytical methods aimed to determine only the unchanged compound as a target analyte, and other bioanalytical methods for 11 compounds contained one to four metabolites as simultaneous target analytes. Stable

Table II. Summary of Bioanalytical Methods Used for Inter-laboratory Cross-validation Studies

Name	Internal standard	Pretreatment method		LC mode	Ionization polarity	Calibration range (ng/mL)	
		Plasma ^a	Urine			Plasma ^a	Urine
Compound A	SIL-IS	PPT	LLE	Reverse phase	Positive	1–1,000	5–5,000
Metabolite A-1	SIL-IS					0.1–100	5–5,000
Compound B	SIL-IS	SPE	SPE	Reverse phase	Positive	0.5–500	10–5,000
Metabolite B-1	SIL-IS					5–500	100–10,000
Metabolite B-2	SIL-IS					0.5–500	10–5,000
Metabolite B-3	SIL-IS					5–500	100–10,000
Metabolite B-4	Structural analogue	LLE	LLE	Reverse phase	Negative	0.5–1,000	10–5,000
Compound C	SIL-IS	LLE	SPE	Reverse phase	Positive	0.01–50	0.5–1,000
Compound D	Structural analogue	SPE	SPE	Reverse phase	Positive	1–2,500	20–10,000
Metabolite D-1	Structural analogue					1–2,500	20–10,000
Metabolite D-2	Structural analogue					1–2,500	20–10,000
Metabolite D-3	Structural analogue					1–2,500	20–10,000
Compound E	SIL-IS	SPE	Dilution	Reverse phase	Positive	0.5–2,000	20–20,000
Metabolite E-1	SIL-IS					0.5–2,000	20–20,000
Compound F	SIL-IS	SPE	Dilution	Reverse phase	Positive	0.1–200	5–10,000
Metabolite F-1	SIL-IS					0.1–200	5–10,000
Compound G	SIL-IS	PPT	Dilution	Reverse phase, UHPLC	Positive	0.1–100	1–1,000
Metabolite G-1	SIL-IS					1–1,000	10–2,000
Metabolite G-2	SIL-IS					1–1,000	10–2,000
Metabolite G-3	SIL-IS					0.1–100	10–2,000
Metabolite G-4	SIL-IS					1–1,000	10–2,000
Compound H	SIL-IS	SPE	SPE	Reverse phase	Positive	0.05–20	0.5–50
Metabolite H-1	SIL-IS					0.5–200	5–500
Metabolite H-2	Structural analogue					0.5–200	5–500
Metabolite H-3	Structural analogue					0.5–200	5–500
Metabolite H-4	Structural analogue					0.5–200	5–500
Compound I	SIL-IS	LLE	LLE	Reverse phase	Positive	0.02–100	0.5–1,000
Compound J	SIL-IS	SPE	SPE	Reverse phase, UHPLC	Positive	0.5–200	20–8,000
Metabolite J-1	SIL-IS					0.5–200	20–8,000
Metabolite J-2	SIL-IS					0.5–200	20–8,000
Metabolite J-3	SIL-IS					0.5–200	20–8,000
Compound K	SIL-IS	LLE	LLE	Reverse phase	Positive	0.1–100	5–5,000
Compound L	Structural analogue	SPE	SPE	Reverse phase, UHPLC	Positive	0.005–10	0.05–100
Compound M	SIL-IS	SPE	SPE	Reverse phase	Positive	0.1–200	5–10,000
Compound N	SIL-IS	PPT	Dilution	Reverse phase, UHPLC	Negative	5–10,000	5–10,000
Metabolite N-1	SIL-IS					5–10,000	5–10,000
Compound O	SIL-IS	SPE	SPE	Strong cation exchange	Positive	0.08–80	10–2,500
Metabolite O-1	Structural analogue					0.2–200	25–6,250
Metabolite O-2	SIL-IS					0.04–40	10–2,500
Compound P	SIL-IS	SPE	SPE	Strong cation exchange	Positive	0.4–400	10–2,500
Metabolite P-1	Structural analogue					2.8–2,800	25–6,250

LC liquid chromatography, SIL-IS stable isotope-labeled internal standard, PPT protein precipitation, SPE solid phase extraction, LLE liquid–liquid extraction, UHPLC ultra-high-pressure liquid chromatography

^a Serum for compound H and its metabolites

isotope-labeled compounds were employed for 30 out of 41 target analytes as internal standards for LC/MS/MS analyses. The pretreatment procedure used for plasma and urine analyses was solid phase extraction, liquid–liquid extraction, or protein precipitation. Simple dilution was also employed for a few of urine methods. Reverse phase liquid chromatography was dominant as the LC mode, and ultra-high-performance liquid chromatography technology was employed for four methods. The dynamic range established for plasma and urine methods was 100- to 5,000-fold.

Evaluation of Inter-laboratory Cross-validation Studies Under the EMA Criterion

The results of inter-laboratory cross-validation studies for 16 compounds and their metabolites in human plasma and urine were evaluated based on the acceptance criterion proposed by the EMA Guideline on BMV (6): The mean accuracy by the different methods should be within 15%. The summary of the evaluation is presented in Table III. The results were evaluated from three different viewpoints: study, analyte, and concentration levels. One study consisted of one

Table III. Summary of Inter-laboratory Cross-validation Studies Under the EMA Criterion

Level of evaluation	Matrices	Number of cross-validation study results (percentage to total, %)			
		Original laboratories		Receiving laboratories	
		Passed	Failed	Passed	Failed
Study	Plasma	15 (93.8)	1 (6.3)	12 (75.0)	4 (25.0)
	Urine	15 (93.8)	1 (6.3)	11 (68.8)	5 (31.3)
	Total	30 (93.8)	2 (6.3)	23 (71.9)	9 (28.1)
Analyte	Plasma	40 (97.6)	1 (2.4)	37 (90.2)	4 (9.8)
	Urine	40 (97.6)	1 (2.4)	34 (82.9)	7 (17.1)
	Total	80 (97.6)	2 (2.4)	71 (86.6)	11 (13.4)
Concentration level	Plasma	122 (99.2)	1 (0.8)	118 (95.9)	5 (4.1)
	Urine	122 (99.2)	1 (0.8)	113 (91.9)	10 (8.1)
	Total	244 (99.2)	2 (0.8)	231 (93.9)	15 (6.1)

to five analytes, and each analyte consisted of three concentration levels. It was found that the majority of results (<90%) passed the acceptance criterion at both the original and receiving laboratories from the evaluation of the concentration level, whereas the number of failures was larger at the receiving laboratories. At the receiving laboratories, only 4.1 and 8.1% of the study results from plasma and urine analyses, respectively, failed to meet the criterion from the concentration evaluation, and these resulted in 25.0 and 31.3% of failure at the study level evaluation. The larger bias (>15% of

nominal) of cross-validation samples was observed regardless of sample pretreatment methods or type of internal standards employed.

Analytical Bias and Variation of Cross-validation Samples

The histograms of 246 of the mean accuracy of the cross-validation samples at the original and receiving laboratories are shown in Fig. 1 containing the mean, SD, and CV of the mean accuracy. Since the observed histograms at both

Fig. 1. Histograms of mean accuracy of cross-validation samples analyzed at **a** original and **b** receiving laboratories

bioanalytical laboratories followed almost a normal distribution, further statistical data analysis was performed on the data set. While the calculated mean of the mean accuracy was very close to 100% at both analytical sites, the SD were 4.8 and 8.1% at both the original and receiving laboratories, respectively, which were significantly different by the *F*-test ($P < 0.01$).

The mean of |mean RE| and mean CV of the cross-validation samples as well as the mean of |mean RE| of the QC samples were calculated and are shown in Fig. 2. Correlation of the |mean RE| and CV of the cross-validation samples are plotted in Fig. 3. While the mean of the |mean RE| of the cross-validation samples was larger at the receiving laboratories, interestingly, there was no obvious difference in the mean CV. In addition, no obvious correlation was observed between the |mean RE| and CV at both laboratories. The mean of the |mean RE| of the cross-validation samples was larger than that of the QC samples at the receiving laboratories while no difference was observed at the original laboratories. The mean of the |mean RE| of QC samples at receiving laboratories was larger than that at original laboratories.

Deviation Between the Cross-validation Samples and QC Samples

The mean accuracy of the cross-validation samples was compared with the individual accuracy of the QC samples with statistical comparison of the analytical variation which

Fig. 2. **a** Mean of |mean RE| of cross-validation samples (X) and QC samples at original and receiving laboratories. **b** Mean CV of cross-validation samples at original and receiving laboratories

Fig. 3. Correlation of |mean RE| and CV of cross-validation samples at **a** original and **b** receiving laboratories

was calculated as the 99% range of the accuracy. The summary of the comparison of the accuracy between the cross-validation samples and QC samples is shown in Table IV. The percentages of both of the QC samples which were outside the 99% range of the accuracy at the low, medium, and high concentration levels were 7.4, 14.8, and 11.1%, respectively, at the original laboratory, and 31.7, 29.3, and 25.6%, respectively, at the receiving laboratory. These results indicated that statistically inexplicable deviations between the accuracy of the cross-validation samples and the QC samples were observed at the receiving laboratory with higher frequency at all the concentration levels evaluated.

The comparisons of the normalized accuracy of the cross-validation samples with the normalized accuracy of the individual QC samples for the selected cross-validation study results where both of the QC samples were out of the 99% range of accuracy at either original or receiving laboratory are depicted in Fig. 4. For visualization purpose, the 99% range and accuracy of cross-validation samples and QC samples were normalized. At all concentration levels, overall, more frequent and larger deviations were observed at the receiving laboratory. On the other hand, it seemed that no obvious difference was observed in terms of the analytical variability represented by error bars between original and receiving laboratories. The direction of the deviation in the QC samples from the normalized mean accuracy of the cross-

Table IV. Summary of Comparison on Accuracy Between Cross-validation Samples and QC Samples

Number of QC samples out of 99% range	Number of cross-validation study results (percentage to total, %)					
	Original laboratories			Receiving laboratories		
	Concentration level			Concentration level		
	Low	Medium	High	Low	Medium	High
0	63 (77.8)	52 (64.2)	52 (64.2)	43 (52.4)	40 (48.8)	42 (51.2)
1	12 (14.8)	17 (21.0)	20 (24.7)	13 (15.9)	18 (22.0)	19 (23.2)
2	6 (7.4)	12 (14.8)	9 (11.1)	26 (31.7)	24 (29.3)	21 (25.6)

QC quality control

validation samples was found to be random through all the concentration levels. The frequency of deviation at the receiving laboratories was similar across the concentration levels.

The mean accuracy of cross-validation samples and QC samples at the original and receiving laboratories which may be showing potential preparation errors was selected from the data used for Fig. 4 and shown in Fig. 5. The results for co-analytes were also depicted if any. The data where the mean accuracy of the cross-validation samples at the receiving laboratory was close to that at the original laboratory are presented in Fig. 5a. Typically, the result for compound C showed that the mean accuracy of the cross-validation samples at both the original and receiving laboratories had bias from their nominal concentrations, which simply suggested preparation error of the cross-validation samples at the original laboratories. In Fig. 5b, on the other hand, the data where mean accuracy of the QC samples at the receiving laboratory was close to that of the cross-validation samples and QC samples at the original laboratory are presented. The results for the co-analytes suggested potential preparation errors. Especially for compound O, the deviation was observed for all the co-analytes that were simultaneously spiked into the cross-validation samples regardless of relatively different physicochemical properties.

Derivation of Acceptance Criterion

Based on the confidence interval approach suggested by H. T. Karnes *et al.* (14), the 99% ranges of mean accuracy, namely, 99% confident interval of accuracy, of the cross-validation samples were calculated at the original and receiving laboratories, which were 12.4 and 20.9%, respectively. Since the fixed range approach is simple and widely accepted in regulated bioanalysis field, 20% cutoff value was derived and proposed as a practical acceptance criterion for inter-laboratory cross-validation studies.

DISCUSSION

From the retrospective data analysis of inter-laboratory cross-validation studies on a broad range of chemical characteristics as shown in Table I, it was suggested that the larger number of failures at receiving laboratories (Table III) was due to the wider distribution of accuracy of the cross-validation samples (Fig. 1). When accuracy and precision

were separately evaluated as those could independently affect the concentration data (14), it was indicated that the larger analytical biases of the cross-validation samples analyzed at the receiving laboratories were caused by neither the analytical variation nor analytical biases in the bioanalytical methods used at the receiving laboratories with assuming that the analytical variation of a bioanalytical method could be estimated from the CV of the cross-validation samples (Figs. 2 and 3). Although the larger bias of cross-validation samples could be partially addressed by slightly larger analytical bias of QC samples at receiving laboratories which may have been due to less experience with the methods, the unexplainable bias was still left for cross-validation samples. Similar results were observed for the selected data where the cross-validation studies were found to fail under the EMA criterion (data not shown).

When the deviations between the accuracy of cross-validation samples and QC samples were evaluated, statistically inexplicable deviation were observed at the receiving laboratories with higher frequency (Table IV), and these results agreed with the previously observed larger analytical bias of cross-validation samples at the receiving laboratories compared with that of the QC samples (Fig. 2). Adsorption, stability, and solubility issues are common in the quantitative analysis of biological matrices, which are related to the characteristics of the target analyte. While the adsorption and stability issues are likely to be more severe at low concentration levels, the solubility issue would be obvious at high concentration levels. Across these three types of analyte-specific issues, on the other hand, the presumed direction of the deviation would be similar where the accuracy of the cross-validation samples may be lower than that of the QC samples since these types of issues are likely to have more impact on the cross-validation samples than the QC samples when analyzed at receiving laboratories. Nevertheless, neither a trend in the direction nor a concentration-dependent deviation was observed as the results of further evaluation for the extracted cross-validation study results (Fig. 4), suggesting that these types of analyte-specific issues could not fully address the deviation between the cross-validation samples and the QC samples. Matrix effect and saturation of the instrument were different types of common analyte-specific issues especially for the bioanalytical methods using LC/MS/MS where the direction of the deviation is likely to be random. However, it is known that the matrix effect is particularly relevant with samples at low concentration levels

Fig. 4. Comparison on normalized accuracy of cross-validation samples with normalized accuracy of individual QC samples for selected cross-validation study results at **a** low, **b** medium, and **c** high concentration levels at original laboratories and at **d** low, **e** medium, and **f** high concentration levels at receiving laboratories. Normalized mean accuracy of cross-validation samples $\pm 2.58 \times CV$, QC quality control sample

(15) and saturation of the instrument is more frequently observed at high concentration levels, indicating that these

issues could not fully address the deviations observed between the cross-validation samples and the QC samples,

Under these circumstances, preparation errors of either the cross-validation samples or the QC samples could be a potential source of bias which may cause concentration-independent random deviations between the cross-validation

samples and QC samples (Fig. 5). More realistically, the combination of preparation error and/or other factors like analyte-specific issues, in other words “inter-laboratory variability,” may cause the large deviation observed for the cross-validation samples. The problem of results represented in Fig. 5b is that an effect of inter-laboratory variability on the cross-validation samples could not be detected until the cross-validation samples are actually shipped and analyzed at the receiving laboratory.

Based on a number of actual inter-laboratory cross-validation results, a practical acceptance criterion, 20% cutoff value, was derived and proposed. The 99% range for the original laboratory was close to the 15% that is the threshold widely adopted for accuracy evaluation of QC samples during the full validation of bioanalytical methods, demonstrating the usefulness of confidence interval approach for deriving an acceptance criterion. The 99% range for receiving laboratories suggested the need of a wider acceptance criterion than the 15% cutoff prescribed by the EMA Guideline (6) in order to address the inter-laboratory variability composed of a preparation error and/or other factors on cross-validation sample analyses. Based on the present data analysis, the 20% cutoff value was considered reasonable under the unique experimental situation where additional variability was shown to be generated from the involvement of additional bioanalytical laboratories. By taking the 20% cutoff, the purpose of inter-laboratory cross-validation studies could be focused on verifying whether analytical biases of the cross-validation samples are within a certain range where analytical bias and variation of the methods as well as inter-laboratory variability are taken into consideration. If 15% cutoff was otherwise employed, additional brush up of fully validated bioanalytical methods is likely to be necessary for substantial number of bioanalytical methods, which would be a high burden for bioanalytical community, as the potential source of inter-laboratory variability could vary and, in some cases, not easy to be identified based on our data analysis. Another option may be to ship all the study samples from multiple study sites to one bioanalytical laboratory in place of establishment of inter-laboratory reliability. However, this option would take more time and risk on sample shipment especially across country border, leading to inefficient use of bioanalytical resources. In addition, the multisite study tends to have more variability at least due to inter-site variability (e.g., site conditions, practices, meals, and a blood collection method). In case of multisite clinical studies, additional variability is likely to be incorporated from intrinsic (e.g., demographics and genetic polymorphism) and extrinsic (e.g., environment, diet, and medical practice) ethnic factors. Under this situation, taking 20% accuracy criterion for inter-laboratory cross-validation study was considered acceptable. From the regulatory point of view, the wider cutoff value proposed in this study is not necessarily against the prescription in the EMA Guideline on BMV (6) considering the room left for the criterion on the cross-validation study. In addition, the Guideline on BMV released from MHLW in

Japan (8) which was recently released adopts the 20% cutoff for the inter-laboratory cross-validation study using QC samples. Although further investigation based on a larger and a variety of data set would be certainly necessary to derive a more appropriate cutoff value and to investigate a source of inter-laboratory variability and its countermeasure, it is hoped that the practical and reasonable acceptance criterion derived from this data analysis could help to facilitate the evaluation and establishment of inter-laboratory reliability of bioanalytical methods while still holding a certain validity on the obtained bioanalytical data, without time-consuming and labor-insensitive further modification of the bioanalytical methods, leading to the effective and efficient use of the concentration data from multiple bioanalytical laboratories during the drug research and development.

CONCLUSIONS

Thorough retrospective data analysis revealed that large deviations observed in inter-laboratory cross-validation studies were not fully caused by the analytical variation and bias attributing to the bioanalytical methods. Rather, it was suggested that inter-laboratory variability like potential preparation error of the cross-validation samples could be a source of analytical bias where the direction of the deviation between cross-validation samples and QC samples was random and not concentration-dependent. An acceptance criterion derived by the confidential interval approach from a number of cross-validation results that the mean accuracy should be within 20% of their nominal, which corresponds to one prescribed in the Guideline on BMV from MHLW in Japan (8) and is a little wider than one in EMA Guideline on BMV (6), would be a practical and reasonable cutoff for the inter-laboratory cross-validation study for QC samples using a LC/MS/MS-based bioanalytical method.

It is hoped that the acceptance criterion for inter-laboratory cross-validation studies proposed from this data analysis could facilitate the establishment of inter-laboratory reliability of bioanalytical methods, leading to more effective and efficient use of concentration data obtained from multiple bioanalytical laboratories with a certain validity during drug research and development.

REFERENCES

1. Shah VP, Midha KK, Dighe S, McGilveray IJ, Skelly JP, Yacobi A, *et al.* Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. Conference report. *Eur J Drug Metab Pharmacokinet.* 1991;16(4):249–55.
2. Shah VP, Midha KK, Findlay JW, Hill HM, Hulse JD, McGilveray IJ, *et al.* Bioanalytical method validation—a revisit with a decade of progress. *Pharm Res.* 2000;17(12):1551–7.
3. Viswanathan CT, Bansal S, Booth B, DeStefano AJ, Rose MJ, Sailstad J, *et al.* Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *Pharm Res.* 2007;24(10):1962–73.

4. Food and Drug Administration. Guidance for industry: bioanalytical method validation. U.S. Department of Health and Human Services, FDA, Center for Drug Evaluation and Research; 2001.
5. Food and Drug Administration. *Draft guidance*, guidance for industry: bioanalytical method validation. U.S. Department of Health and Human Services, FDA, Center for Drug Evaluation and Research, Center for Veterinary Medicine; 2013.
6. European Medicines Agency. Guideline on bioanalytical method validation. European Medicines Agency, Committee for Medicinal Products for Human Use; 2011.
7. Yoneyama T, Inoue N, Tachiki H, Togashi K, Nakayama A, Kudo T, *et al.* Draft guideline on bioanalytical method validation in Japan by Japan Bioanalysis Forum. *Pharm Med Device Regul Sci.* 2012;43(8):750–60.
8. Ministry of Health, Labour and Welfare. Guideline on bioanalytical method validation in pharmaceutical development; 2013.
9. Gilbert MT, Barinov-Colligon I, Miksic JR. Cross-validation of bioanalytical methods between laboratories. *J Pharm Biomed Anal.* 1995;13(4–5):385–94.
10. Muirhead DC, Smart TS. Cross validation in bioanalysis: why, when and how? *Chromatographia.* 2000;52:S72–5.
11. Gansser D. Experience with a cross-validation approach. *Chromatographia.* 2002;55:S71–4.
12. Lin ZJ, Li W, Weng N. Capsule review on bioanalytical method transfer: opportunities and challenges for chromatographic methods. *Bioanalysis.* 2011;3(1):57–66.
13. Xu X, Ji QC, Jemal M, Gleason C, Shen JX, Stouffer B, *et al.* Fit-for-purpose bioanalytical cross-validation for LC-MS/MS assays in clinical studies. *Bioanalysis.* 2013;5(1):83–90.
14. Karnes HT, Shiu G, Shah VP. Validation of bioanalytical methods. *Pharm Res.* 1991;8(4):421–6.
15. Trufelli H, Palma P, Famiglini G, Cappiello A. An overview of matrix effects in liquid chromatography-mass spectrometry. *Mass Spectrom Rev.* 2011;30(3):491–509.