7.0 Reliability of the *In Vitro* Pyrogen Test Methods

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Repeatability refers to the closeness of agreement among test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM 1997, 2003). Intra-laboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Inter-laboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes a quantitative and/or qualitative analysis of intralaboratory repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, positive), where applicable.

An evaluation of intralaboratory repeatability and reproducibility could be conducted because *in vitro* pyrogen test data were available from replicate wells within individual experiments, and from replicate experiments within the individual laboratories. In addition, comparable data were available from each of the three laboratories that performed the validation studies, which allowed an evaluation of interlaboratory reproducibility.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of *In Vitro* Pyrogen Test Methods

The quality of a reliability evaluation depends on the extent to which the substances tested adequately represent the range of physicochemical characteristics and response levels that the test method should be capable of evaluating. The rationale for selecting the substances used in the validation studies was discussed in **Section 3.1**. In response to the ICCVAM PWG request for data on other relevant test materials (e.g., medical devices, biologics, etc.) with these test methods, ECVAM summarized published and unpublished studies on snake venom sera, medical devices, dialysate, and lipidic formulations (see question #3 in **Appendix B**).

Each sample contained the appropriate endotoxin spike concentration when tested at its Maximum Valid Dilution (MVD). The MVD takes into account the endotoxin limit concentration (ELC) and the detection limit of the particular test method. The U.S. and European Pharmacopeias assign ELCs for drugs based on their specific administered dose, route of administration, and dosing regimen. Based on the selected threshold pyrogen dose of 0.5 EU/mL (see **Section 4.0**), and the decision criteria used in the validation studies to identify a pyrogenic response (≥0.5 EU/mL, see **Section 5.0**), a concentration of 0.5 EU/mL was used as the detection limit for the *in vitro* test methods when calculating the MVDs for each of the test substances.

7.2 Analysis of Intralaboratory Repeatability and Reproducibility

Intralaboratory repeatability analyses were performed using the OD values obtained for each test with each spiked sample. All analyses of intra- and inter-laboratory reproducibility were performed on the classifications of pyrogenic or non-pyrogenic, rather than on the absolute

OD values generated in each run. Analyses of intra-laboratory reliability include a CV analysis for the log-transformed OD_{450} measurements, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. 1996). According to Section 7.2 of each ECVAM BRD, the analyses focused on the CV because existing data has demonstrated that there is a direct relationship between the mean responses and the variation (e.g., empirical variance or standard deviation). Moreover, the CV should be distributed symmetrically around a constant factor if the mean-variance relationship is linear.

7.2.1 Intralaboratory Repeatability

In the ECVAM validation study, intralaboratory repeatability of each test method was evaluated by testing saline and various endotoxin spikes (0.06 to 0.5 EU/mL) in saline and evaluating the closeness of agreement among OD readings for cytokine measurements at each concentration. Each experiment was conducted up to three times for each test method. Up to 20 replicates per concentration were tested and results indicated that variability in OD measurements increased with increasing endotoxin concentration, but the variability was not so great as to interfere with distinguishing the 0.5 EU/mL spike concentration (i.e., the threshold for pyrogenicity) from the lower concentrations. **Table 7-1** details the study design for each of these evaluations. With the exception of the Cryo WB/IL-1β test method, at least four different study designs were employed for each test method. Appendix C of the ECVAM Cryo WB/IL-1β BRD (see **Appendix A**) indicates that because intralaboratory reliability was extensively evaluated in the WB/IL-1β test method, only a subset (n=2) of these studies was conducted as part of a "catch-up validation" study. Based on the "acceptable" intralaboratory performance in this subset of studies, additional studies were not considered necessary.

With regard to plate-to-plate variation, the ECVAM Trial Data Report (see **Appendix C**) states that the data obtained from each ELISA plate (i.e., 96-well format) must be considered as a whole and cannot be compared to other ELISA plates due to uncontrollable variation. Therefore, it was recommended that each ELISA plate should include all controls (e.g., negative control, positive control, negative product control, and positive product control) required for the analytical procedure.

Table 7-1 Intralaboratory Repeatability Assessed with Saline Spiked with WHO-LPS 94/580

Experiment	Study Design	Test Method				
		MM6/IL-6	PBMC/IL-6	WB/IL-1β	WB/IL-6	Cryo WB/IL-1β[1]
1A	Endotoxin concentration (EU/mL)	0, 0.25, 0.5	0, 0.25, 0.5	0, 0.5	0, 0.5	0, 0.5
	N (per spike)	20	20	32	20	32
	Repetitions of experiment	1	1	1	1	1
1B	Endotoxin concentration (EU/mL)	0, 0.063, 0.125, 0.25, 0.5	0, 0.063, 0.125, 0.25, 0.5	0, 0.063, 0.125, 0.25, 0.5	0, 0.063, 0.125, 0.25, 0.5	0, 0.063, 0.125, 0.25, 0.5
	N (per spike)	12	12	12	10	12
	Repetitions of experiment	1	1	1	1	1
2A	Endotoxin concentration (EU/mL)	0, 0.25, 0.5	0, 0.5	0, 0.5	0, 0.25, 0.5	ND
	N (per spike)	20	8	12	8	ND
	Repetitions of experiment	3	3	3	3	ND
2B	Endotoxin concentration (EU/mL)	0, 0.25, 0.5	0, 0.063, 0.125, 0.25, 0.5	0, 0.25, 0.5	0, 0.5	ND
	N (per spike)	20	8	8	5	ND
	Repetitions of experiment	3	3	3	8	ND
2C	Endotoxin concentration (EU/mL)	ND	0, 0.125, 0.25, 0.5	0, 0.5	ND	ND
	N (per spike)	ND	8	5	ND	ND
	Repetitions of experiment	ND	8	8	ND	ND

Abbreviations: Cryo = Cryopreserved; EU/mL = Endotoxin units/mL; IL = Interleukin; LPS = Lipopolysaccharide; MM6 = Mono Mac 6; N = number of replicates; ND = Not done; PBMC = Peripheral blood mononuclear cells; WB = Whole blood; WHO = World Health Organization

[1]The Cryo WB/IL-1β test method was included in a catch-up validation study to assess intralaboratory reliability in a subset of experiments (n=2).

7.2.2 Intralaboratory Reproducibility

Intralaboratory reproducibility was evaluated using three marketed pharmaceuticals spiked with various concentrations of endotoxin (see **Table 3-2**). Three identical, independent runs were conducted in each of the three testing laboratories, with the exception of the Cryo WB/IL-1β test method.[5] The correlations (expressed as a percentage of agreement) between

[5] The ECVAM Cryo WB/IL-1 test method BRD states that there was no direct assessment of intralaboratory reproducibility because such an evaluation was performed in the WB/IL-1 test method, and the authors assumed that variability would not be affected by the use of cryopreserved blood.

pairs of the independent runs (i.e., run 1 vs. run 2; run 1 vs. run 3; run 2 vs. run 3) were determined and the mean of these three values was calculated. In all reproducibility analyses, a single run consisted of each of the substances assayed in quadruplicate. Acceptability criteria for each run included a CV analysis to remove highly variable responses from the analyses. The criterion used to identify outliers ranged from CV <0.25 to CV <0.45, depending on the method being considered, and was arbitrarily set based on results using saline spiked with endotoxin. As an example, for the MM6/IL-6 test method, the CV for any single spike concentration was ≤ 0.12, and therefore, the outlier criterion was set at 0.25.

Agreement between different runs was determined for each substance in three laboratories. As shown in **Table 7-2**, the agreement across three runs in an individual lab ranged from 75% to 100%

.

Table 7-2 Intralaboratory Reproducibility of *In Vitro* Pyrogen Test Methods

Run Comparison[1]	WB/IL-1β			Cryo WB/IL-1β			WB/IL-6			PBMC/IL-6			MM6/IL-6		
	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
1 vs 2	92% (11/12)	100% (8/8)	100% (12/12)	ND[3]	ND	ND	75% (9/12)	92% (11/12)	100% (12/12)	92% (11/12)	100% (12/12)	100% (12/12)	100% (12/12)	92% (11/12)	100% (12/12)
1 vs 3	83% (10/12)	88% (7/8)	92% (11/12)	ND	ND	ND	100% (12/12)	92% (11/12)	100% (12/12)	100% (12/12)	100% (12/12)	92% (11/12)	100% (12/12)	92% (11/12)	92% (11/12)
2 vs 3	92% (11/12)	NI[4]	92% (11/12)	ND	ND	ND	75% (9/12)	92% (11/12)	100% (12/12)	92% (11/12)	100% (12/12)	92% (11/12)	100% (12/12)	100% (12/12)	92% (11/12)
Mean	89%	n.c.	95%	ND	ND	ND	83%	92%	100%	95%	100%	95%	100%	95%	95%
Agreement[2] across 3 runs	83%	n.c.	92%	ND	ND	ND	75%	92%	100%	92%	100%	94%	100%	92%	92%

Abbreviations: Cryo = Cryopreserved; IL= Interleukin; MM6 = Mono Mac 6; n.c. = Not calculated; ND = Not done; NI = Not included; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Comparison among 3 individual runs within each laboratory

[2]All possible combinations of runs among the 3 laboratories were compared.

[3]Not done. The ECVAM Cryo WB/IL-1β BRD states that an assessment of intralaboratory reproducibility was performed using the WB IL-1β (fresh blood) test method, and it was assumed that intralaboratory variability would not be affected by the change to cryopreserved blood assayed in 96-well plates.

[4]Not included due to lack of sufficient data. The sensitivity criteria were not met for 1 of 3 substances in run 2, and 1 of 3 substances in run 3.

[This Page Intentionally Left Blank]

7.2.3 Interlaboratory Reproducibility

Interlaboratory reproducibility was evaluated in two different studies. In both studies, each run from one laboratory was compared with all runs of another laboratory. The proportions of similarly classified samples provide a measure of reproducibility. In the first study, the interlaboratory reproducibility was evaluated using results from three marketed pharmaceuticals spiked with endotoxin and tested in triplicate in each of the three laboratories. As shown in **Table 7-3**, the agreement across three laboratories for each test method (where three runs per laboratory were conducted) ranged from 58% to 86%, depending on the test method considered. In comparison, the agreement across three laboratories for the Cryo WB/IL-1β test method, for which only one run per laboratory was conducted, was 92%.

Table 7-3 Interlaboratory Reproducibility of *In Vitro* Pyrogen Test Methods

Lab Comparison[1]	Agreement Between Laboratories[1]				
	WB/IL-1β (Tube)	Cryo WB/IL-1β	WB/IL-6	PBMC/IL-6	MM6/IL-6
1 vs 2	92% (77/84)[2]	92% (11/12)[3]	72% (78/108)	81% (87/108)	97% (105/108)
1 vs 3	77% (83/108)	92% (11/12)[3]	75% (81/108)	86% (93/108)	89% (96/108)
2 vs 3	68% (57/84)[2]	92% (11/12)[3]	97% (105/108)	89% (96/108)	86% (93/108)
Mean	79%	92%	81%	85%	90%
Agreement across 3 labs[4]	58% (167/288)[2]	92% (11/12)[3]	72% (234/324)	78% (252/324)	86% (279/324)

Abbreviations: Cryo = Cryopreserved; IL= Interleukin; MM6 = Mono Mac 6; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Data from three substances (see **Table 3-2**) spiked with endotoxin (WHO-LPS 94/580 [*E. coli* O113:H10:K-]) at 0, 0.5, and 1.0 endotoin units/mL (EU/mL), with 0 EU/mL tested in duplicate, were tested three times in three different laboratories, with the exception of Cryo WB/IL-1β (only the preliminary run from each laboratory used for analysis).

[2]Some of the runs did not meet the assay acceptance criteria and therefore were excluded from the analysis.

[3]For the Cryo WB/IL-1β test method, each substance tested only once in each laboratory.

[4]All possible combinations of runs among the 3 laboratories were compared (with the exception of Cryo WB/IL-1β, which was only tested once in each laboratory, resulting in only one possible combination per substance).

In the second study, interlaboratory reproducibility was evaluated with the same 10 substances used for evaluating accuracy. In this study, each of the substances was spiked with four concentrations of endotoxin (with one conentration spiked in replicate) and tested once in each of three laboratories. As shown in **Table 7-4**, the agreement across three laboratories for each test method ranged from 57% to 88%, depending on the test method considered. The extent and order of agreement among laboratories was the same for both studies; the WB/IL-1β tube method showed the least agreement (57-58%) and the Cryo WB/IL-1β test method showed the most (88-92%).

Table 7-4 Interlaboratory Reproducibility of *In Vitro* Pyrogen Test Methods

Lab Comparison[1]	Agreement Between Laboratories[1]						
	WB/IL-1β (Tube)	WB/IL-1β (Plate)	Cryo WB/IL-1β	WB/IL-6	PBMC/IL-6	PBMC/IL-6 (Cryo)	MM6/IL-6
1 vs 2	73% (35/48)	88% (37/42)	84% (38/45)	85% (41/48)	84% (42/50)	96% (48/50)	90% (45/50)
1 vs 3	82% (40/49)	90% (35/39)	88% (21/24)	85% (41/48)	86% (43/50)	76% (38/50)	90% (43/48)
2 vs 3	70% (33/47)	92% (43/47)	100% (25/25)	88% (44/50)	90% (45/50)	80% (40/50)	83% (40/48)
Mean	75%	90%	91%	86%	87%	84%	88%
Agreement across 3 labs	57% (27/47)	85% (33/39)	88% (21/24)	79% (38/48)	80% (40/50)	76% (38/50)	81% (39/48)

Abbreviations: Cryo = Cryopreserved; IL = Interleukin; MM6 = Mono Mac 6; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Data from 10 substances spiked with endotoxin (WHO-LPS 94/580 [*E. coli* O113:H10:K-]) at 0, 0.25, 0.5, and 1.0 endotoxin units/mL (EU/mL), with 0.5 EU/mL tested in duplicate, were tested once in three different laboratories.

7.3 Historical Positive and Negative Control Data

No historical control data were provided for any of the five *in vitro* pyrogen test methods. However, the intralaboratory repeatability analysis described in **Section 7.2.1** included repeat testing of both spiked (0.5 EU/mL endotoxin) and non-spiked saline, and the accumulated positive and negative control values, respectively for each of the methods. As a result, the database that was accumulated during the ECVAM validation studies provides an indication of the range and variability in responses for the positive and negative controls.