3.0 Validation Status of *In Vitro* Pyrogen Test Methods

The following is a synopsis of the information in the ICCVAM BRD, which reviews the available data and information for each of the five test methods. The ICCVAM BRD describes the current validation status of the five *in vitro* pyrogen test methods, including what is known about their reliability and accuracy, the scope of the substances tested, and standardized protocols used for the validation study. The ICCVAM BRD may be obtained electronically from the NICEATM/ICCVAM website (http://iccvam.niehs.nih.gov/) or by contacting NICEATM via email at niceatm@niehs.nih.gov. A hard copy of the ICCVAM BRD may be requested by email or by mail to NICEATM, NIEHS, P.O. Box 12233, Mail Drop EC-17, Research Triangle Park, NC 27709.

3.1 Test Method Description

According to the ECVAM submission, these *in vitro* pyrogen test methods are intended for the detection of Gram-negative endotoxin contained in substances intended for parenteral use (e.g., pharmaceuticals, biologics, medical devices). These methods are based on the detection of the release of proinflammatory cytokines (i.e., IL-1β or IL-6) from human monocytes or monocytoid cells induced by exposure to a product contaminated with Gram-negative endotoxin.

3.1.1 General Test Method Procedures

The *in vitro* pyrogen test methods measure cytokine release from monocytes or monocytoid cells (i.e., WB, PBMCs, or the MM6 cell line) by using an enzyme-linked immunosorbent assay (ELISA) that includes monoclonal or polyclonal antibodies specific for either IL-1β or IL-6. The amount of endotoxin present is determined by comparing the values of endotoxin equivalents produced by WB cells exposed to the test substance to those exposed to an internationally harmonized Reference Standard Endotoxin (RSE)[4] or an equivalent standard expressed in Endotoxin Units (EU)/mL. A product is considered to be pyrogenic if the endotoxin concentration exceeds the Endotoxin Limit Concentration (ELC) for the test substance.

3.1.2 Protocol Similarities and Differences

Although there are differences among the five *in vitro* pyrogen test methods, the basic procedural steps are consistent across all test methods:

- The test substance is mixed with a suspension of human-derived cells.
- The mix of cells and test product is incubated for a specific time.
- The concentration of pro-inflammatory cytokines (e.g., IL-1β, IL-6) is measured with an ELISA by comparison to a standard curve.

[4]RSEs are internationally harmonized reference standards (e.g., WHO-lipopolysaccharide [LPS] 94/580 *Escherichia coli* [*E. coli*] O113:H10:K-; U.S. Pharmacopeia [USP] RSE *E. coli* LPS Lot G3E069; USP RSE *E. coli* Lot G; FDA *E. coli* Lot EC6). Equivalent endotoxins include commercially available *E. coli*-derived LPS Control Standard Endotoxin or other *E. coli* LPS preparations that have been calibrated with an appropriate RSE.

- The endotoxin content is calculated by comparing the measured concentration of pro-inflammatory cytokines to an endotoxin standard curve.
- A test substance is considered pyrogenic if the estimated endotoxin concentration of the test substance exceeds the ELC for the test substance.

3.2 Validation Database

The test substances selected for use in the validation studies were marketed parenteral pharmaceuticals. No biological or medical device products were included in the validation study. A total of 13 test substances were included in the performance analysis of each of the five *in vitro* test methods. Ten substances (**Table 3-1**), each spiked with four concentrations of endotoxin (0, 0.25, 0.5, and 1.0 EU/mL, with 0.5 EU/mL tested in duplicate), were used to evaluate accuracy. Three substances (**Table 3-2**), each spiked with three concentrations of endotoxin (0, 0.5, and 1.0 EU/mL, with 0 EU/mL tested in duplicate), were used to assess intralaboratory reproducibility. Interlaboratory reproducibility was evaluated in two different studies. The first study tested the substances listed in **Table 3-2** in triplicate in each of three laboratories. In the second study, interlaboratory reproducibility was tested using the substances in **Table 3-1**, which were tested once in each of three laboratories.

Table 3-1 Parenteral Drugs Used in the Validation Studies for Determining Test Method Accuracy[1]

Test Substance[2]	Active Ingredient	Source	Lot Number(s)	Indication	MVD (-fold)
Beloc®	Metoprolol tartrate	Astra Zeneca	DA419A1	Heart dysfunction	140
Binotal®	Ampicillin	Grünenthal	117EL2	Antibiotic	140
Ethanol 95%	Ethanol	B. Braun	2465Z01	Diluent	35
Fenistil®	Dimetindenmaleat	Novartis	21402 26803[3]	Antiallergic	175
Glucose 5%	Glucose	Eifelfango	1162 3132[3]	Nutrition	70
MCP®	Metoclopramid	Hexal	21JX22	Antiemetic	350
Orasthin®	Oxytocin	Hoechst	W015	Initiation of delivery	700
Sostril®	Ranitidine	Glaxo Wellcome	1L585B 3H01N[3]	Antiacidic	140
Syntocinon®	Oxytocin	Novartis	S00400	Initiation of delivery	-
Drug A - 0.9% NaCl	0.9% NaCl	-	-	-	35
Drug B - 0.9% NaCl	0.9% NaCl	-	-	-	70

Abbreviations: MVD = Maximum valid dilution

[1]Each substance was tested in all five *in vitro* pyrogen test methods.

[2]Each test substance was spiked with 0, 0.25, 0.5, or 1.0 Endotoxin Units (EU)/mL of endotoxin (WHO-LPS 94/580 [*E. coli* O113:H10:K-]), with 0.5 EU/mL tested in duplicate. Each sample contained the appropriate spike concentration when tested at its MVD.

[3]Indicates the lot numbers used in the catch-up validation study for the Cryopreserved whole blood/Interleukin-1β test method.

Table 3-2 Parenteral Drugs Used in the Validation Studies for Determining Test Method Reproducibility[1]

Test Substance[2]	Source	Agent	Indication
Gelafundin®	Braun Melsungen	Gelatin	Transfusion
Haemate®	Aventis	Factor VIII	Hemophilia
Jonosteril®	Fresenius	Electrolytes	Infusion

[1]Each substance was tested in all five *in vitro* pyrogen test methods.
[2]Each test substance was spiked with 0, 0.5, or 1.0 Endotoxin Units (EU)/mL of endotoxin (WHO-LPS 94/580 [*E. coli* O113:H10:K-]), with 0 EU/mL tested in duplicate. Each sample contained the appropriate spike concentration when tested at its maximum valid dilution.

3.3 Reference Test Method Data

The historical RPT studies were conducted at the Paul Ehrlich Institut (PEI), which supports regional German regulatory authorities, provides marketing approval of certain marketed biological products (e.g., sera, vaccines, test allergens), and functions as a World Health Organization (WHO) collaborating center for quality assurance of blood products and *in vitro* diagnostics. The unit for pyrogen and endotoxin testing of the PEI is accredited following the International Organization for Standardization (ISO) and the International Electrotechnical Commission 17025 (ISO 2005). In a request for additional information from ECVAM, it was stated that the RPT data was generated according to the European Pharmacopeia (EP) monograph, but the detailed protocol used by this laboratory was not provided.

These data were generated for internal quality control studies from 171 rabbits (Chinchilla Bastards). Chinchilla Bastards are reported to be a more sensitive strain than the New Zealand White rabbit strain for pyrogenicity testing (Hoffmann et al. 2005b). However, the USP (USP 2007) and the EP (EP 2005) do not prescribe a specific rabbit strain for the RPT.

3.4 Test Method Accuracy

The ability of the *in vitro* pyrogen test methods to correctly identify the presence of Gram-negative endotoxin was evaluated using parenteral pharmaceuticals spiked with endotoxin (WHO-LPS 94/580 [*E. coli* O113:H10:K-]). As described in **Section 3.2**, 10 substances (see **Table 3-1**) spiked with four concentrations of endotoxin (i.e., 0, 0.25, 0.5, or 1.0 EU/mL, with 0.5 EU/mL tested in duplicate) were used for the evaluation. The individual spike concentrations in each substance were tested once, using each test method, in three different laboratories, providing a total of 150 runs (i.e., 10 substances x 5 spike solutions x 3 laboratories = 150). Outliers were identified using Dixon's test (i.e., significance level of α = 0.01) and subsequently excluded from the evaluation, which resulted in fewer than a total of 150 runs per evaluation (Dixon 1950; Barnett et al. 1984). A comparison of the results for the *in vitro* test methods indicates that the number of runs excluded was greatest for the Cryo WB/IL-1β and WB/IL-1β (plate method) test methods, which had 30 and 11 runs excluded, respectively. No other test method had more than three runs excluded.

As described in **Section 3.3**, no RPTs were conducted in parallel with the *in vitro* pyrogen test methods during the ECVAM validation studies. Instead, historical RPT data from rabbits tested with endotoxin were used to establish a threshold pyrogen dose (i.e., the endotoxin

dose at which fever was induced in 50% of the rabbits). This historical data were subsequently used to establish the limit of detection (i.e., 0.5 EU/mL) that the *in vitro* test methods being validated must meet. Accordingly, the *in vitro* call was compared to the "true status" (based on the known endotoxin spike concentration) of the sample. The resulting calls were used to construct 2x2 contingency tables, which were used to calculate the resulting test method performance values.

The accuracy of each *in vitro* pyrogen test method for correctly identifying samples spiked with 0.5 or 1.0 EU/mL endotoxin as positive and samples spiked with 0 or 0.25 EU/mL endotoxin as negative was evaluated. As provided in **Table 3-3**, accuracy ranged from 81% to 93%, sensitivity ranged from 73% to 99%, specificity ranged from 77% to 97%, false negative rates ranged from 1% to 27%, and false positive rates ranged from 3% to 23%.

Table 3-3 Accuracy of *In Vitro* Pyrogen Test Methods[1]

Test Method	Accuracy[2]	Sensitivity[3]	Specificity[4]	False Negative Rate[5]	False Positive Rate[6]
Cryo WB/IL-1β	92% (110/120)	97% (75/77)	81% (35/43)	3% (2/77)	19% (8/43)
MM6/IL-6	93% (138/148)	96% (85/89)	90% (53/59)	5% (4/89)	10% (6/59)
PBMC/IL-6	93% (140/150)	92% (83/90)	95% (57/60)	8% (7/90)	5% (3/60)
PBMC/IL-6 (Cryo)[7]	87% (130/150)	93% (84/90)	77% (46/60)	7% (6/90)	23% (14/60)
WB/IL-6	92% (136/148)	89% (79/89)	97% (57/59)	11% (10/89)	3% (2/59)
WB/IL-1β (Tube)	81% (119/147)	73% (64/88)	93% (55/59)	27% (24/88)	7% (4/59)
WB/IL-1β (96-well plate)[8]	93% (129/139)	99% (83/84)	84% (46/55)	1% (1/84)	16% (9/55)

Abbreviations: Cryo = Cryopreserved; EU/mL = Endotoxin units per milliliter; IL = Interleukin; MM6 = Mono Mac 6; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Data shown as a percentage (number of correct runs/total number of runs), based on results of 10 parenteral drugs tested in each of three different laboratories. Samples of each drug were tested with or without being spiked with a Gram-negative endotoxin standard (0, 0.25, 0.5, or 1.0 EU/mL, with 0.5 EU/mL tested in duplicate).

[2]Accuracy = the proportion of correct outcomes (positive and negative) of a test method.

[3]Sensitivity = the proportion of all positive substances that are classified as positive.

[4]Specificity = the proportion of all negative substances that are classified as negative.

[5]False negative rate = the proportion of all positive substances that are falsely identified as negative.

[6]False positive rate = the proportion of all negative substances that are falsely identified as positive.

[7]A modification of the PBMC/IL-6 test method that uses Cryo PBMCs.

[8]A modification of the WB/IL-1β test method that uses 96-well plates instead of tubes for the test substance incubation.

3.5 Test Method Reliability

Intralaboratory repeatability was evaluated by testing saline spiked with various concentrations of endotoxin (0, 0.06, 0.125, 0.25, 0.5, and 1.0 EU/mL) and then evaluating the closeness of agreement among OD readings for cytokine measurements at each concentration. For each test method, each experiment was conducted up to three times. From 5 to 32 replicates per concentration were tested and results indicated that variability in OD measurements increased with increasing endotoxin concentration. However, the variability

did not interfere with distinguishing the 0.5 EU/mL spike concentration (i.e., the threshold for pyrogenicity) from the lower concentrations.

Intralaboratory reproducibility was evaluated using three marketed pharmaceuticals spiked with three concentrations of endotoxin (i.e., 0, 0.5, and 1.0 EU/mL, with 0 EU/mL tested in duplicate). 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 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 products assayed in quadruplicate. Acceptability criteria for each run included a Coefficient of Variation (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 3-4**, the agreement across three runs in an individual lab ranged from 75% to 100%.

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 3-5**, 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 (excludes the Cryo WB/IL-1β test method, which used only one run per laboratory). However, if the WB/IL-1β tube method is excluded, the range of agreement across laboratories is 72% to 86%. 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%.

---

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

[This Page Intentionally Left Blank]

Table 3-4 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%	NC	95%	ND	ND	ND	83%	92%	100%	95%	100%	95%	100%	95%	95%
Agreement[2] across 3 runs	83%	NC	92%	ND	ND	ND	75%	92%	100%	92%	100%	92%	100%	92%	92%

Abbreviations: Cryo = Cryopreserved; IL = Interleukin; MM6 = Mono Mac 6; NC = 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]

Table 3-5 Interlaboratory Reproducibility of *In Vitro* Pyrogen Test Methods: Study One

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 EU/mL, with 0 EU/mL spiked 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 (0, 0.25, 0.5, and 1.0 EU/mL, with 0.5 EU/mL spiked in duplicate) and tested once in each of three laboratories. As shown in **Table 3-6**, 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 test method showed the least agreement (57-58%) and the Cryo WB/IL-1β test method showed the most (88-92%).

Table 3-6 Interlaboratory Reproducibility of *In Vitro* Pyrogen Test Methods: Study Two

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 EU/mL, with 0.5 EU/mL spiked in duplicate, were tested once in three different laboratories.

3.6 Animal Welfare Considerations: Reduction, Refinement, and Replacement

The currently accepted pyrogen test methods require the use of rabbits or horseshoe crab hemolymph. The proposed *in vitro* pyrogen test methods use monocytoid cells of human origin, obtained either from WB donations or from an immortalized cell line. The capability of these five *in vitro* assays to detect Gram-negative endotoxin suggests that they may reduce or eventually replace the use of rabbits and/or horseshoe crab hemolymph for pyrogen testing. However, at the present time, the RPT detects classes of pyrogens that have neither been examined nor validated with the *in vitro* pyrogen test methods and thus, the RPT will still be required for most test substances.

Human blood donations are required for four of the five *in vitro* test methods (WB/IL-1β, WB/IL-6, Cryo WB/IL-1β, and PBMC/IL-6) proposed as replacements for the RPT, and as such, no animals will be used when these assays are appropriate for use. While the collection of human blood is a common medical procedure, the many aspects of human blood collection must be considered to ensure that human donors are treated appropriately, and that such collection and use is in accordance with all applicable regulations, policies, and guidelines.