6.0 Relevance of the *In Vitro* Pyrogen Test Methods

6.1 Accuracy of *In Vitro* Pyrogen Test Methods

A critical component of an ICCVAM evaluation of the validation status of a test method is an assessment of its relevance. The measure of relevance used in this evaluation is the performance of the new test in identifying pyrogens as compared to the performance of the current reference method (ICCVAM 2003). This aspect of assay performance is typically evaluated by calculating:

- Accuracy (also referred to as concordance): the proportion of correct outcomes (positive and negative) of a test method
- Sensitivity: the proportion of true positive substances that are correctly classified as positive
- Specificity: the proportion of true negative substances that are correctly classified as negative
- Positive predictivity: the proportion of correct positive responses among substances testing positive
- Negative predictivity: the proportion of correct negative responses among substances testing negative
- False positive rate: the proportion of true negative substances that are falsely identified as positive
- False negative rate: the proportion of true positive substances that are falsely identified as negative

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 (with one concentration 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). The quality criteria outlined in **Table 2-1** were used to identify outliers. These outliers were subsequently excluded from the evaluation, which resulted in less than a total of 150 runs per evaluation.

As described in **Section 4.2**, 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.

6.1.1 Relevance of the Cryo WB/IL-1β Test Method

Of the 150 available runs for the Cryo WB/IL-1β test method, 10 runs showed excessive variability but no significant outliers among the four replicates (i.e., CV >45%) resulting in their exclusion from the analysis. An additional 20 runs (from one of the three participating laboratories) did not qualify according to one or more of the criteria outlined in **Table 2-1**. Therefore, a total of 120 runs were used in the performance analysis which showed that the Cryo WB/IL-1β test method has an accuracy of 92% (110/120), a sensitivity of 97% (75/77), a specificity of 81% (35/43), a false negative rate of 3% (2/77), and a false positive rate of 19% (8/43) (see **Table 6-1**).

Table 6-1 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.

6.1.2 Relevance of the MM6/IL-6 Test Method

Of the 150 available runs for the MM6/IL-6 test method, two showed excessive variability among the four replicates (i.e., CV >25%), resulting in their exclusion from the analysis. No runs were excluded based on the criteria outlined in **Table 2-1**. Therefore, a total of 148 runs was used in the performance analysis. Based on this analysis, the MM6/IL-6 test method has an accuracy of 93% (138/148), a sensitivity of 96% (85/89), a specificity of 90% (53/59), a false negative rate of 4% (4/89), and a false positive rate of 10% (6/59) (see **Table 6-1**).

6.1.3 Relevance of the PBMC/IL-6 Test Method

None of the 150 available runs for the PBMC/IL-6 test method showed excessive variability (i.e., CV >40%) and all runs met the criteria outlined in **Table 2-1**. Therefore, all 150 runs were included in the performance analysis. Based on this analysis, the PBMC/IL-6 test method has an accuracy of 93% (140/150), a sensitivity of 92% (83/90), a specificity of 95% (57/60), a false negative rate of 8% (7/90), and a false positive rate of 5% (3/60) (see **Table 6-1**).

6.1.3.1 Relevance of the PBMC/IL-6 Method When Using Cryo PBMCs

As indicated in **Table 2-1**, the PBMC/IL-6 test method protocol was also conducted using a modified protocol that included Cryo PBMCs. None of the 150 available runs for this modification of the PBMC/IL-6 test method showed excessive variability (i.e., CV >40%) and all runs met the criteria outlined in **Table 2-1**. Therefore, all runs were included in a performance analysis. Based on this analysis, the PBMC/IL-6 test method, when using Cryo PBMCs, has an accuracy of 87% (130/150), a sensitivity of 93% (84/90), a specificity of 77% (46/60), a false negative rate of 7% (6/90), and a false positive rate of 23% (14/60). The high false positive rate can be attributed to a large number of false positives (50% [10/20]) in one of the three laboratories (the false positive rate in the remaining two laboratories is 10%).

6.1.4 Relevance of the WB/IL-6 Test Method

None of the 150 available runs for the WB/IL-6 test method showed excessive variability (i.e., CV >45%) and all runs met the criteria outlined in **Table 2-1**. However, two samples were mishandled by one of the testing laboratories, and thus the two associated runs were excluded from the analysis. As a result, 148 runs were included in the performance analysis for the detection of Gram-negative endotoxin. Based on this analysis, the WB/IL-6 test method has an accuracy of 92% (136/148), a sensitivity of 89% (79/89), a specificity of 97% (57/59), a false negative rate of 11% (10/89), and a false positive rate of 3% (2/59) (see **Table 6-1**).

6.1.5 Relevance of the WB/IL-1β Test Method

Of the 150 available runs for the WB/IL-1β test method, three showed excessive variability among the four replicates (i.e., CV >45%), resulting in their exclusion from the analysis. No runs were excluded based on the criteria outlined in **Table 2-1**. Therefore, a total of 147 runs was used in the performance analysis. Based on this analysis, the WB/IL-1β test method has an accuracy of 81% (119/147), a sensitivity of 73% (64/88), a specificity of 93% (55/59), an false negative rate of 27% (24/88), and a false positive rate of 7% (4/59) (see **Table 6-1**). Improved performance statistics for the WB/IL-1β test method associated with the use of 96-well plates is summarized below (**Section 6.1.5.1**).

6.1.5.1 Relevance of the WB/IL-1β Test Method When Using 96-Well Plates

As indicated in **Table 2-1**, the WB/IL-1β test method protocol was also conducted using a modified protocol that used 96-well plates instead of individual tubes. Of the 150 available runs for this modification of the WB/IL-1β test method, 11 showed excessive variability (i.e., CV >45%). No runs were excluded based on the criteria outlined in **Table 2-1**. Therefore, a

total of 139 runs were included in a performance analysis. Based on this analysis, the WB/IL-1β test method, when using 96-well plates, has an accuracy of 93% (129/139), a sensitivity of 99% (83/84), a specificity of 84% (46/55), a false negative rate of 1% (1/84), and a false positive rate of 16% (9/55).

6.2 Summary of the Performance Statistics for *In Vitro* Pyrogen Test Methods

The performance of the *in vitro* pyrogen test methods for the detection of Gram-negative endotoxin (based on 10 parenteral pharmaceuticals, each spiked with four concentrations of endotoxin, with one spiked in duplicate) was evaluated. As outlined in **Table 6-1**, this analysis indicated that the accuracy among the test methods ranged from 81% to 93%, sensitivity ranged from 89% to 99%, specificity ranged from 81% to 97%, false negative rates ranged from 1% to 27%, and false positive rates ranged from 3% to 23%.

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.

6.2.1 Discordant Results

It was not possible to make a direct comparison between the RPT and *in vitro* pyrogen test results without the availability of parallel testing data (i.e., same test substance tested using the *in vitro* and *in vivo* methods). Therefore, *in vitro* results that are discordant from the RPT could not be identified with these studies. Discordant results reflect either a failure of the *in vitro* test method to identify Gram-negative endotoxin (i.e., false negative) when spiked into a test substance at 0.5 EU/mL (i.e., the threshold concentration established based on historical data from the RPT) or 1.0 EU/mL, or to incorrectly indicate the presence of Gram-negative endotoxin (i.e., false positive) when spiked into a test substance at 0 or 0.25 EU/mL. As shown in **Table 6-2**, false positive rates ranged from 7% to 47% when spiked into a test substance at 0.25 EU/mL and from 0% to 3% when spiked with 0 EU/mL. Similarly, false negative rates ranged from 2% to 39% when spiked into a test substance at 0.5 EU/mL and from 0% to 3% when spiked with 1.0 EU/mL.

6.2.2 Strengths and Limitations of In Vitro Pyrogen Test Methods

The limitations of these test methods have not been fully explored and identified. As described in **Section 3.0**, the substances tested do not adequately represent the range of products that are tested with these methods. For this reason, pre-testing product specific validation will be necessary to establish if a particular test substance/material is appropriate for evaluation using these *in vitro* test methods. A recognized limitation of the *in vitro* methods is the lack of data to determine their responses to, and suitability for, non-endotoxin pyrogens that can be detected by the RPT. Additional limitations of these test methods are outlined in the ECVAM response to ICCVAM PWG questions (see question #4 in **Appendix B**). However, an advantage to these *in vitro* test methods is that they are derived from human tissues, and thus avoid potential uncertainty associated with cross-species extrapolation.

Table 6-2 Predictivity of *In Vitro* Pyrogen Test Methods for Each Endotoxin Spike Concentration[1]

Test Method	Endotoxin Spike Concentration								Overall Totals	
	Negative for Pyrogen (< 0.5 EU/mL)				Positive for Pyrogen (≥ 0.5 EU/mL)					
	0 EU/mL		0.25 EU/mL		0.5 EU/mL		1.0 EU/mL		False Negative	False Positive
	Correct	False Positive[2]	Correct	False Positive	False Negative[3]	Correct	False Negative	Correct		
Cryo WB/IL-1β	100% (24/24)	0% (0/24)	58% (11/19)	42% (8/19)	4% (2/51)	96% (49/51)	0% (0/26)	100% (26/26)	3% (2/77)	19% (8/43)
MM6/IL-6	100% (30/30)	0% (0/30)	79% (23/29)	17% (6/29)	7% (4/59)	93% (55/59)	0% (0/30)	100% (30/30)	5% (4/89)	10% (6/59)
PBMC/IL-6	100% (30/30)	0% (0/30)	90% (27/30)	10% (3/30)	12% (7/60)	88% (53/60)	0% (0/30)	100% (30/30)	8% (7/90)	5% (3/60)
PBMC/IL-6 (Cryo)[4]	100% (30/30)	0% (0/30)	53% (16/30)	47% (14/30)	10% (6/60)	90% (54/60)	0% (0/30)	100% (30/30)	7% (6/90)	23% (14/60)
WB/IL-6	100% (30/30)	0% (0/30)	93% (27/29)	7% (2/29)	17% (10/59)	83% (49/59)	0% (0/30)	100% (30/30)	11% (10/89)	3% (2/59)
WB/IL-1β (Tube)	97% (28/29)	3% (1/29)	90% (27/30)	10% (3/30)	39% (23/59)	61% (36/59)	3% (1/29)	97% (28/29)	27% (24/88)	7% (4/59)
WB/IL-1β (96-well plate)[5]	100% (28/28)	0% (0/28)	67% (18/27)	33% (9/27)	2% (1/55)	98% (54/55)	0% (0/29)	100% (29/29)	1% (1/84)	16% (9/55)

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

[1]Data shown as a percentage (number of correct, false positive, or false negative 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]False positive rate = the proportion of all negative substances that are falsely identified as positive.

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

[4]A modification of the PBMC/IL-6 test method using cryopreserved PBMCs.

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