ICCVAM Background Review Document:

Validation Status of Five *In Vitro* Test Methods Proposed for Assessing Potential Pyrogenicity of Pharmaceuticals and Other Products

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

May 2008
NIH Publication No. 08-6391

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

This document is available electronically at:
http://iccvam.niehs.nih.gov/methods/pyrogen/pyrogen.htm

Table of Contents

List of Tables

*Tables 1 and 6-1 are identical. Table 1 provides supporting information for the Executive Summary, while Table 6-1 provides the information in Section 6.1, where test method accuracy is discussed in greater detail.

List of Abbreviations and Acronyms

AWIPT	Adsorb, Wash, *In Vitro* Pyrogen Test
BET	Bacterial Endotoxin Test
BRD	Background Review Document
CBER	Center for Biologics Evaluation and Research
CCAC	Canadian Council on Animal Care
CDER	Center for Drug Evaluation and Research
CDRH	Center for Devices and Radiological Health
CEC	Commission of the European Communities
CFR	Code of Federal Regulations
CLB	Central Laboratory for the Blood Transfusion Service
Cryo	Cryopreserved
CSF	Colony Stimulating Factor
CV	Coefficient of variation
CVM	Center for Veterinary Medicine
DMSO	Dimethyl sulfoxide
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures)
ECVAM	European Centre for the Validation of Alternative Methods
EDQM	European Directorate for the Quality of Medicines
EEC	European Economic Community
ELC	Endotoxin limit concentration
ELISA	Enzyme-linked immunosorbent assay
EMEA	European Medicines Agency
EP	European Pharmacopeia
EPA	U.S. Environmental Protection Agency
ESAC	ECVAM Scientific Advisory Committee
EU	European Union
EU/mL	Endotoxin units/mL
FDA	U.S. Food and Drug Administration
Fn	Fibronectin
FR	*Federal Register*
GLP	Good Laboratory Practice
GM-CSF	Granulocyte-Macrophage-Colony Stimulating Factor
Hb	Hemoglobin
hGH	Human growth hormone
HSA	Human serum albumin
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IFN-γ	Interferon-γ
IgG	Immunoglobulin G
IL	Interleukin
i.m.	Intramuscular
ISO	International Standards Organization
IU	International units

i.v.	Intravenous
JP	Japanese Pharmacopeia
LAL	*Limulus* Amebocyte Lysate
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
MEM	Minimum essential medium
MM6	Mono Mac 6
MVD	Maximum valid dilution
NA	Not applicable
NC	Negative control
n.c.	Not calculated
ND	Not done
n.d.	Not detectable
NF	National Formulary
NI	Not included
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NPC	Negative product control
OECD	Organisation for Economic Co-operation and Development
OD	Optical density
PBMC	Peripheral Blood Mononuclear Cell
PC	Positive control
PEI	Paul Ehrlich Institut
PFS	Pyrogen-free saline
PG	Peptidoglycan
PHA	Phytohemagglutinin
PPC	Positive product control
PWG	Pyrogenicity Working Group
QA	Quality assurance
rec	Recombinant
RPT	Rabbit Pyrogen Test
SD	Standard deviation
SOP	Standard operating procedure
TBE	Tick-borne encephalitis
THP-1	Acute monocyte leukemia cell line
TLR	Toll-like Receptor
TNF-α	Tumor Necrosis Factor-α
U	Units
U.K.	United Kingdom
U.S.	United States
U.S.C.	United States Code
USDA	U.S. Department of Agriculture
USP	U.S. Pharmacopeia
USPTO	United States Patent and Trademark Office
UV	Ultraviolet
WB	Whole blood

WHO	World Health Organization
x *g*	Times gravity

Interagency Coordinating Committee on the Validation of Alternative Methods: Agency Representatives

Agency for Toxic Substances and Disease Registry
★Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
★ Marilyn L. Wind, Ph.D. **(Chair)**
◊ Kristina Hatlelid, Ph.D.
Joanna Matheson, Ph.D.

Department of Agriculture
★ Jodie Kulpa-Eddy, D.V.M. **(Vice-Chair)**
◊ Elizabeth Goldentyer, D.V.M.

Department of Defense
★ Robert E. Foster, Ph.D.
◊ Patty Decot
Peter J. Schultheiss, D.V.M., D.A.C.L.A.M.
Harry Salem, Ph.D.

Department of Energy
★ Michael Kuperberg, Ph.D.
◊ Marvin Stodolsky, Ph.D.

Department of the Interior
★ Barnett A. Rattner, Ph.D.
◊ Sarah Gerould, Ph.D.

Department of Transportation
★ George Cushmac, Ph.D.
◊ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Science Coordination and Policy
★ Karen Hamernik, Ph.D.
Office of Research and Development
◊ Julian Preston, Ph.D.
Suzanne McMaster, Ph.D.
Office of Pesticides Programs
Amy Rispin, Ph.D.
Deborah McCall
OECD Test Guidelines Program
Jerry Smrchek, Ph.D.

Food and Drug Administration
Office of Science
★ Suzanne Fitzpatrick, Ph.D., D.A.B.T.
Center for Drug Evaluation and Research
◊ Abigail C. Jacobs, Ph.D.
Paul C. Brown, Ph.D.
Center for Devices and Radiological Health
Melvin E. Stratmeyer, Ph.D.
Center for Biologics Evaluation and Research
Richard McFarland, Ph.D., M.D.
Ying Huang, Ph.D.
Center for Food Safety and Nutrition
David G. Hattan, Ph.D.
Robert L. Bronaugh, Ph.D.
Center for Veterinary Medicine
Devaraya Jagannath, Ph.D.
M. Cecilia Aguila, D.V.M.
National Center for Toxicological Research
William T. Allaben, Ph.D.
Paul Howard, Ph.D.
Office of Regulatory Affairs
Lawrence A. D'Hoostelaere, Ph.D.

National Cancer Institute
★ Alan Poland, M.D.
◊ T. Kevin Howcroft, Ph.D.

National Institute of Environmental Health Sciences
★ William S. Stokes, D.V.M., D.A.C.L.A.M.
◊ Raymond R. Tice, Ph.D.
Rajendra S. Chhabra, Ph.D., D.A.B.T
Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
★ Paul Nicolaysen, V.M.D.
◊ K. Murali Rao, MD, Ph.D.

National Institutes of Health
★ Margaret D. Snyder, Ph.D.

National Library of Medicine
◊ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
★ Surender Ahir, Ph.D

★ Principal Agency Representative
◊ Alternate Principal Agency Representative

Acknowledgements

The following individuals are acknowledged for their contributions to the ICCVAM in vitro pyrogen test method evaluation process.

Pyrogenicity Working Group (PWG)

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

U.S. Food and Drug Administration
Center for Biologics Evaluation and Research
Mustafa Akkoyunlu, M.D., Ph.D.
Pankaj Amin
Christine Anderson
Kimberly Benton, Ph.D.
Joseph George, Ph.D.
Christopher Joneckis, Ph.D.
Richard McFarland, Ph.D, M.D. (Chair)
Center for Drug Evaluation and Research
David Hussong, Ph.D.
Abigail Jacobs, Ph.D.
Robert Mello, Ph.D.
Amy Rosenberg, M.D.
Daniela Verthelyi, M.D., Ph.D.
Jiaqin Yao, Ph.D.
Center for Devices and Radiological Health
Raju Kammula, D.V.M., Ph.D., D.A.B.T.
Ramesh Panguluri
Center for Food Safety and Nutrition
Penelope Rice, Ph.D.

U.S. Environmental Protection Agency
Office of Science Coordination and Policy
Karen Hamernik, Ph.D.

National Institute of Environmental Health Sciences
William Stokes, D.V.M., D.A.C.L.A.M.
Raymond Tice, Ph.D.

U.S. Department of Agriculture
Jodie Kulpa-Eddy, D.V.M.

Liaison, European Centre for the Validation of Alternative Methods
Marlies Halder, Ph.D.[1]

[1]Dr. Halder was responsible for the trial plan of the catch-up validation study, but she was not involved in the evaluation of the data.

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences

William Stokes, D.V.M., D.A.C.L.A.M.	Director; Project Officer
Raymond Tice, Ph.D.	Deputy Director
Deborah McCarley	Special Assistant; Asst. Project Officer

NICEATM Support Contract Staff (Integrated Laboratory Systems [ILS], Inc.)

David Allen, Ph.D.	Principal Investigator
Douglas Winters, M.S.	Project Manager
Neepa Choksi, Ph.D.	Senior Staff Toxicologist
Judy Strickland, Ph.D., D.A.B.T.	Senior Staff Toxicologist
Frank Deal, M.S.	Staff Toxicologist
Elizabeth Lipscomb, Ph.D.	Staff Toxicologist
Thomas Burns, M.S.	Senior Project Coordinator/Technical Writer
Michael Paris	Senior Project Coordinator/Technical Writer
Patricia Ceger, M.S.	Project Coordinator/Technical Writer
James Truax, M.A.	Project Coordinator/Technical Writer
Catherine Sprankle	Communications Specialist/Web Developer
Linda Litchfield	Meeting Planner and Coordinator

Consultants for ILS, Inc.

Joe Haseman, Ph.D.	Statistical Consultant
Errol Zeiger, Ph.D., J.D.	Editorial Consultant

Preface

Endotoxin, a bacterial pyrogen also known as lipopolysaccharide, is an integral component of the Gram-negative bacterial cell membrane. Endotoxin directly interacts with host monocytoid cells to induce the release of a variety of proinflammatory cytokines (e.g., interleukin [IL]-1β, IL-6, tumor necrosis factor-α). In addition to an initial febrile reaction, excessive release of these cytokines during Gram-negative bacterial sepsis can lead to multiple organ failure and death. For this reason, it is critical that parenteral pharmaceuticals, fluids for injection, medical devices, and human biological products be properly and accurately evaluated for the presence of endotoxin prior to their clinical or veterinary use. The original pyrogen test, the rabbit pyrogen test (RPT), was developed in 1941 to limit to an acceptable level the risks of febrile reaction in the patient following administration of, or contact with, the product of concern. While the RPT continues to serve this purpose well, an endotoxin test using a hemolymph extract (i.e., "blood") from the horseshoe crab (i.e., the bacterial endotoxin test [BET]) was developed in the early 1970's as an *in vitro* alternative to the RPT for the detection of Gram-negative endotoxin. In 1980, the United States (U.S.) Food and Drug Administration (FDA) published guidelines for use of the BET as an end-product test for human and animal drug products. The U.S., European, and Japanese Pharmacopeias currently recognize both test methods for pyrogen testing (i.e., RPT and BET). The BET is recognized for its sensitivity to the presence of endotoxins from Gram-negative bacteria, but it also has some limitations, including its inability to respond to non-endotoxin pyrogens, as well as its susceptibility to interference from certain types of materials (e.g., products with high protein and lipid levels, glucans). In contrast, the RPT is capable of detecting both endotoxin and non-endotoxin pyrogens.

More recent efforts have focused on the development of *in vitro* test systems that might achieve or exceed the sensitivity of the BET and the RPT. Test systems based on the activation of human monocytes *in vitro* have been developed that take advantage of the role of these cells in the fever response. The European Centre for the Validation of Alternative Methods (ECVAM), a unit of the Institute for Health and Consumer Protection at the European Commission's Joint Research Centre, conducted a validation study to independently evaluate the usefulness of six *in vitro* pyrogen test methods. The study was financed by the European Commission within the 5th Framework Programme of Directorate General Research and was recently published (Hoffmann et al. 2005a). Since two tests based on the acute monocyte leukemia cell line THP-1 did not meet the validation criteria, they are not included in the peer review. In 2004, the University of Konstanz (Germany) carried out catch-up validation studies of two tests using Cryopreserved whole blood (Cryo WB/IL-1β) or blood cells (cryopreserved or fresh peripheral blood mononuclear cells [PBMC]/IL-6), the results of which were recently published (Schindler et al. 2006).

Based on these studies, in June 2005, ECVAM submitted background review documents (BRDs) for five of these test methods, which were proposed as replacements for the RPT, to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). The five test methods are:

- The Human Whole Blood (WB)/IL-1β *In Vitro* Pyrogen Test

- The Human WB/IL-1β *In Vitro* Pyrogen Test: Application of Cryo Human WB
- The Human WB/IL-6 *In Vitro* Pyrogen Test
- The Human PBMC/IL-6 *In Vitro* Pyrogen Test
- The Monocytoid Cell Line Mono Mac 6/IL-6 *In Vitro* Pyrogen Test

For simplicity, the submitted studies are referred to collectively as the ECVAM validation study in this document.

ICCVAM, which is charged with coordinating the technical evaluations of new, revised, and alternative test methods with regulatory applicability (ICCVAM Authorization Act of 2000, [42 U.S. Code 285*l*-3, available at http://iccvam.niehs.nih.gov/docs/about_docs/PL106545.htm]), unanimously agreed that the five submitted *in vitro* test methods should have a high priority for evaluation. An ICCVAM Pyrogenicity Working Group (PWG) was established to work with the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to carry out these evaluations. The PWG consists of knowledgeable scientists from ICCVAM member agencies. The PWG functions included reviewing draft test method BRDs, recommending proposed performance standards, identifying and recommending scientists for independent peer review panels, preparing questions for expert or peer review Panels, developing ICCVAM draft test method recommendations regarding the usefulness and applicability of the alternative test methods for regulatory testing, and recommending necessary future validation studies. ICCVAM and NICEATM also collaborate closely with ECVAM. Accordingly, an ECVAM liaison was designated for the ICCVAM PWG to provide additional clarification and information during the evaluation and review process.

NICEATM, which administers the ICCVAM and provides scientific support for ICCVAM activities, subsequently prepared a comprehensive draft BRD containing all of the information and data from the validation studies for each of the five *in vitro* test methods. A request for any other data and information on these test methods was made through a 2005 *Federal Register* (*FR*) request (Vol. 70, No. 241, pp. 74833-74834, December 16, 2005; available at http://ntp-apps.niehs.nih.gov/iccvampb/searchFR.cfm), through the ICCVAM electronic mailing list, and through direct requests to over 100 interested stakeholders. No additional data or information was submitted in response to this request.

The draft BRD was made publicly available on the NICEATM-ICCVAM website (http://iccvam.niehs.nih.gov). Comments from the public and scientific community were welcomed and were provided to the Panel and made available on the NICEATM-ICCVAM website (see *FR* notice [Vol. 71, No. 238, pp. 74533-74534, December 12, 2006], available at http://iccvam.niehs.nih.gov).

The independent review of the usefulness and limitations of the five test methods took place in a public meeting of the independent peer review panel (Panel) on February 6, 2007 at the National Institutes of Health in Bethesda, Maryland. The Panel considered the information and data available in the draft BRD. The Panel's independent peer review report was then made available for public comment on the NICEATM-ICCVAM website (see *FR* notice [Vol. 72, No. 89, pp. 26395-26396, May 9, 2007], available at http://iccvam.niehs.nih.gov). Following the Panel meeting, ICCVAM and the PWG considered the Panel's report and

public comments, and prepared this final BRD. ICCVAM and the PWG also considered the Panel's report, comments from the public and from the Scientific Advisory Committee on Alternative Toxicological Methods, and information in this BRD, and prepared final test method recommendations that will be provided to U.S. Federal agencies and made available to the public. These final recommendations are included in the ICCVAM Test Method Evaluation Report, which is available at http://iccvam.niehs.nih.gov/methods/pyrogen/pyrogen.htm, in accordance with the ICCVAM Authorization Act of 2000.

We acknowledge the ECVAM scientists who participated in the management of the validation studies and who prepared the ECVAM BRDs. We especially acknowledge Dr. Marlies Halder, ECVAM Liason to the PWG, for valuable information and comments throughout the review process. The efforts of many individuals who contributed to the preparation of the ICCVAM BRD are also gratefully acknowledged. These include Drs. David Allen and Elizabeth Lipscomb, Bradley Blackard, Catherine Sprankle, James Truax, and Doug Winters of Integrated Laboratory Systems, Inc., the NICEATM support contractor, as well as the members of the ICCVAM PWG and ICCVAM representatives who subsequently reviewed and provided comments throughout the process leading to this final version. We also want to thank Dr. Raymond Tice, Deputy Director of NICEATM, for his coordination efforts for this project. Finally, we want to recognize the excellent leadership of the PWG Chair, Dr. Richard McFarland, FDA.

William S. Stokes, D.V.M., D.A.C.L.A.M.
Rear Admiral, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Marilyn Wind, Ph.D.
U.S. Consumer Product Safety Commission
Chairman, ICCVAM

[This Page Intentionally Left Blank]

EXECUTIVE SUMMARY

This Background Review Document (BRD), prepared by the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM), provides a comprehensive description and analyses of the data and information supporting the validity of five *in vitro* pyrogen test methods. This BRD includes data from previously conducted validation studies and from previously published and unpublished data. The test methods are:

- The Human Whole Blood (WB)/Interleukin (IL)-1β *In Vitro* Pyrogen Test
- The Human WB/IL-1β *In Vitro* Pyrogen Test: Application of Cryopreserved (Cryo) Human WB
- The Human WB/IL-6 *In Vitro* Pyrogen Test
- The Human Peripheral Blood Mononuclear Cell (PBMC)/IL-6 *In Vitro* Pyrogen Test
- The Monocytoid Cell Line Mono Mac 6 (MM6)/IL-6 *In Vitro* Pyrogen Test

The validation studies evaluated the test methods for their ability to detect the presence of Gram-negative endotoxin that had been spiked into a range of injectable pharmaceuticals. This ICCVAM BRD provides information and data that support the current validation status of the *in vitro* pyrogen test methods. It discusses what is known about their relevance[1] and reliability[2], the types of substances tested, and the standardized test method protocols used to generate data for each test method.

Information in this ICCVAM BRD is based on data from five individual BRDs submitted by the European Centre for the Validation of Alternative Methods (ECVAM), a unit of the Institute for Health and Consumer Protection at the European Commission's Joint Research Centre (see **Appendix A**), to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. The ECVAM BRDs were prepared according to the ICCVAM submission guidelines (ICCVAM 2003). The ECVAM BRDs will also help agencies to assess whether the proposed test methods are acceptable for regulatory applications. Each ECVAM BRD summarizes the validation studies conducted for an individual *in vitro* pyrogen test method. This ICCVAM BRD compares and contrasts the performance of these five test methods.

This ICCVAM BRD also summarizes information from published studies and additional unpublished data provided by ECVAM. **Section 9.0** of this document discusses *in vitro* pyrogen test method studies that could not be included in the performance analyses because appropriate study details, test method results, or *in vivo* rabbit pyrogen test (RPT) reference data were not available. An online literature search for additional data on the proposed *in vitro* pyrogen test methods identified nineteen studies that contained relevant data. ECVAM also provided additional unpublished data in response to a request for additional information related to the validation studies (**Appendices B** and **C**).

---

[1]Relevance is the extent to which a test method correctly predicts or measures an effect, and includes the "accuracy" or "concordance" of the method.

[2]Reliability is a measure of how well a test method can be reproduced at different times and in different laboratories. It is assessed by calculating reproducibility both within and among laboratories and repeatability within laboratories.

An independent peer review panel (Panel) assessed the ICCVAM BRD for completeness and any errors or omissions. The Panel also evaluated the validation status of the proposed test methods in the ICCVAM BRD.

The *in vitro* pyrogen test methods discussed in this BRD measure release of the pro-inflammatory cytokines IL-1β or IL-6 in response to exposure to Gram-negative endotoxin. The test methods use monocytoid cells contained in WB, isolated PBMCs, or the MM6 cell line. No data were provided from the validation studies supporting the usefulness of these test methods for pyrogens other than endotoxins.

ICCVAM surveyed regulatory agencies in the United States (U.S.) to determine whether any of the proposed *in vitro* test methods have been considered for regulatory use where submission of test data is required. Regulatory practice in the U.S. and in the European Union is to accept pyrogen test method data for a specific product after the test method has been validated for that specific product. The ECVAM BRDs note that the U.S. Food and Drug Administration has accepted data from the PBMC test developed by Novartis and Baxter Healthcare. In this instance, the PBMC test results were used in conjunction with RPT and Bacterial Endotoxin Test data to support the safety testing of a single specific drug product (New Drug Application Number 16-267/S-037).

The predominant difference between the *in vitro* pyrogen test methods is the type of cells used. The following basic steps are consistent among all methods:

- Interference testing is performed to verify that a test substance does not interfere with either the cell system used or with the specific cytokine-specific enzyme-linked immunosorbent assay (ELISA).
- The test substance is mixed with a suspension of human-derived blood cells.
- The concentration of of the specific proinflammatory cytokine (e.g., IL-1β, IL-6) is measured using an ELISA, and is compared to the response curve of an endotoxin standard.
- An internationally accepted endotoxin standard (World Health Organization-lipopolysaccharide [WHO-LPS] 94/580 *Escherichia coli* [*E. coli*] O113:H10:K-), or an endotoxin standard that has been calibrated against this standard, is used to generate the standard response curve for the assay. The endotoxin activity of a test substance is calculated by comparing the induced cytokine release with that induced by the endotoxin standard.
- A product "passes" (i.e., is considered negative for endotoxin pyrogen activity) if the cytokine response to the test substance is less than that induced by 0.5 endotoxin units/mL (EU/mL).

The ability of the *in vitro* pyrogen test methods to correctly identify the presence of Gram-negative endotoxin was evaluated using 10 parenteral pharmaceuticals spiked with endotoxin (WHO-LPS 94/580 *E. coli* O113:H10:K-). Each drug, spiked with four concentrations of endotoxin, was tested once in three different laboratories. As indicated in **Table 1**, analysis of the five *in vitro* test methods indicated that accuracy among the test methods ranged from 81% to 93%, sensitivity ranged from 89% to 99%, specificity ranged

from 89% to 99%, specificity ranged from 81% to 97%, false negative rates[3] ranged from 1% to 27%, and false positive rates[4] ranged from 3% to 23%.

Table 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.

The RPT and *in vitro* pyrogen test results can be compared if the same substance is tested using both the *in vivo* RPT and *in vitro* methods (i.e., parallel testing data). However, because no RPT data were generated with the same test samples used in the *in vitro* test methods, the accuracy of the *in vitro* test results could not be compared directly with that of the RPT.

The limitations of these five *in vitro* test methods have not been fully evaluated. For this reason, product-specific validation will be necessary to establish if a particular test substance or material is appropriate for evaluation using these *in vitro* test methods. One identified limitation of the *in vitro* test methods is the lack of data to determine their responses to, and suitability for, pyrogens other than endotoxins that are currently detected by the RPT. However, a potential advantage of these *in vitro* test methods is that they are derived from

[3]False negative rates reflect a failure of the *in vitro* test method(s) to identify Gram-negative endotoxin spiked into a test substance at the threshold concentration (0.5 EU/mL) established based on historical data from the RPT.

[4]False positive rates reflect that the *in vitro* test method(s) identified the presence of Gram-negative endotoxin when it was not present.

human tissues, which avoids potential uncertainty associated with cross-species extrapolation.

Repeatability within individual laboratories was determined for each *in vitro* test method, using saline and various endotoxin spikes (0.06 to 0.5 EU/mL) to evaluate the closeness of agreement among optical density (OD) readings for cytokine measurements at each concentration. Up to 20 replicates per concentration were tested, and results indicated that variability in OD measurements increased with increasing endotoxin concentration. However, the variability was low enough that the threshold for pyrogenicity could still be detected (i.e., the 0.5 EU/mL spike concentration could still be distinguished from the lower concentrations).

Reproducibility within individual laboratories was evaluated using three marketed pharmaceuticals spiked with various concentrations of endotoxin. 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 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. Agreement between two runs within a single laboratory ranged from 75% to 100%, with mean values ranging from 83% to 100%. Agreement across three runs within a single laboratory ranged from 75% to 100%.

Reproducibility across all laboratories was evaluated in two different studies in which each run from one laboratory was compared to all other runs of another laboratory. The proportion of equally qualified samples provided a measure of reproducibility. In the first reproducibility study, three marketed pharmaceutical products were spiked with either saline control or various concentrations of endotoxin, and each sample was tested in triplicate in each of three different laboratories, except for Cryo WB/IL-1β. In the catch-up validation study of Cryo WB/IL-1β, each sample was tested once in each laboratory. The agreement across the three laboratories for each test method ranged from 58% to 86%[6], depending on the test method used, and 92% for the Cryo WB/IL-1β test method.

In the second study, reproducibility was determined using the results from the 10 drugs used in the accuracy analysis. Each drug was spiked with four concentrations of endotoxin and tested once in each of three laboratories. The agreement across three laboratories for each test method ranged from 57% to 88%, depending on the test method used. The extent and order of agreement among laboratories were the same for both studies: the WB/IL-1β test method showed the least agreement (57% to 58%), and the Cryo WB/IL-1β test method showed the most (88% to 92%).

This ICCVAM BRD provides a comprehensive summary of available data used to determine the usefulness and limitations of five *in vitro* pyrogen test methods for detecting Gram-negative endotoxin. It discusses what is currently known about their relevance and reliability,

---

[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 (fresh blood) test method, and the authors assumed that variability is not affected by the change to cryopreserved blood.

[6]However, a modification of the WB/IL-1 test method (using 96-well plates for the test substance incubation) resulted in agreement among laboratories of 83% to 92% when tested once in each laboratory.

the types of the substances tested, and the standardized test method protocols used to generate data for each test method. The information in this BRD was used by ICCVAM to finalize its recommendations for test method uses, standardized test method protocols, and future studies to further characterize the usefulness and limitations of these test methods. These test method recommendations will be provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S. Code § 285*l*-2 through 285*l*-5), available at: http://iccvam.niehs.nih.gov/about/about_ICCVAM.htm. Agency responses to ICCVAM will be available on the NICEATM-ICCVAM website (http://iccvam.niehs.nih.gov) 180 days after agency receipt of the recommendations.

[This Page Intentionally Left Blank]

1.0 Introduction And Rationale For The Proposed Use Of *In Vitro* Pyrogen Test Methods

1.1 Introduction

1.1.1 Historical Background of In Vitro Pyrogen Test Methods and the Rationale for Their Development

A brief summary of the historical development of the five *in vitro* pyrogen test methods was provided in Section 1.1.1 of each Background Review Document (BRD) provided by the European Centre for the Validation of Alternative Methods (ECVAM), a unit of the Institute for Health and Consumer Protection at the European Commission's Joint Research Centre. These BRDs were provided to the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and are included in **Appendix A**[1]. This section includes supplementary information and provides a context for United States (U.S.) regulatory considerations.

Pyrogenic substances (i.e., substances that induce fever) may originate from a variety of biological or synthetic/manufacturing sources. They may also be released from microbiological organisms such as bacteria, viruses, and fungi during cell death or following immunological attack (i.e., cell damage or death due to a local or systemic immune response). One of the most potent pyrogenic materials is bacterial endotoxin, which is an outer membrane component of the Gram-negative bacteria cell wall. Pyrogens may also be found in processing and packaging materials, chemicals, raw materials, or equipment used during the manufacturing of parenteral drugs or medical devices. The presence of endotoxins in otherwise sterile biological preparations such as parenteral drugs suggests the presence of past or current bacterial contamination.

The induction of fever by these pyrogenic substances is a complex process and multiple mechanisms are thought to be involved. It is likely that the specific pathway, or combinations of pathways, involved in the production of a fever response depends on a number of variables (e.g., the properties of the pyrogenic substance and the route of administration). In general, pyrogenic substances cause leukocytes (i.e., neutrophils, monocytes/macrophages, and lymphocytes) to release cytokines (e.g., interleukin [IL]-1β, IL-6, and Tumor Necrosis Factor-α [TNF-α]) that act as endogenous pro-inflammatory mediators, often referred to as "endogenous pyrogens" (Dinarello 1999). Once released, these cytokines act on the central nervous system to promote the synthesis of prostaglandins, ultimately producing of a fever response (Dinarello 1999; Netea et al. 2000). These cytokines have been shown to be associated with the fever response induced by pyrogenic substances in both humans and rabbits (Dinarello 1999). Certain bacterial products (e.g., endotoxin) can also stimulate cytokine production directly through the activation of Toll-like receptors (TLRs) (Dinarello 1999; Netea et al. 2000).

[1]References to ECVAM BRD sections are in normal type to distinguish them from references to ICCVAM BRD sections, which are in **boldface** type.

The translation of released cytokines into a fever response is largely mediated by circumventricular organs. These small neuronal cell groups allow neurons to come in contact with a variety of circulating substances directly from the bloodstream, which are thought to control the febrile response through projections to sites in the hypothalamus and brain stem (Saper and Breder 1994; Dinarello 1999; Beutler and Rietschel 2003).

The U.S., European, and Japanese Pharmacopeias currently recognize two test methods for pyrogen testing, the *in vivo* rabbit pyrogen test (RPT) and the *in vitro* bacterial endotoxin test (BET), also referred to as the *Limulus* amebocyte lysate (LAL) test. The BET is accepted because of its sensitivity to the presence of Gram-negative endotoxins. However, the test method has well documented limitations, including its inability to respond to non-endotoxin pyrogens, as well as its susceptibility to interference from certain types of materials (e.g., high protein and lipid levels, glucans). In contrast, the RPT is capable of detecting both endotoxin and non-endotoxin pyrogens. However, disadvantages of the RPT include the need for interspecies extrapolation from rabbits to humans.

In 2002, a total of 243,838 rabbits were used in the U.S. for all research and testing purposes, of which 6,324 rabbits were reported as experiencing more than slight or momentary pain and/or distress where anesthetics, analgesics, or tranquilizers could not be administered for scientific reasons (U.S. Department of Agriculture [USDA] 2002). Eight of these cases were specifically attributed to pyrogenicity testing, presumably based on induction of a fever response (USDA 2002). Thus, although the potential for more than slight or momentary pain and/or distress exists for pyrogenicity testing when a fever response is induced, it does not appear that a fever response is common. In the European Union (EU), approximately 313,000 total rabbits were used for all scientific purposes in 2005 (CEC 2007). Of these, approximately 276,000 rabbits were used for pharmaceutical products and medical device testing (i.e., either research and development, production and quality control, or toxicological and other safety evaluations). Although the number of rabbits specifically used for pyrogenicity testing was not reported, it is likely that this number is significantly less than the total of 276,000. Additional animal use numbers, including data reported from Canada and the United Kingdom (U.K.) are summarized in **Section 10.1**.

An *in vitro* test system that combines the sensitivity of the BET with the wide range of pyrogens detectable by the RPT would be an obvious improvement for pyrogen testing. With this intention, test systems based on the activation of human monocytes *in vitro* were developed that take advantage of an increased understanding of the biological mechanisms responsible for the human fever reaction (Dinarello 1999). Initial efforts focused on peripheral blood mononuclear cells (PBMC), which release proinflammatory cytokines when exposed to endotoxin (Duff and Atkins 1982; Dinarello et al. 1984). A number of similar test systems, using either whole blood (WB), PBMCs, or monocytoid cell lines (e.g., Mono Mac 6 [MM6], THP-1) were subsequently developed (Tsuchiya et al. 1980; Poole et al. 1988; Ziegler-Heitbrock et al. 1988; Hartung and Wendel 1996; Hartung et al. 2001; Poole et al. 2003; Gaines Das et al. 2004). Five test systems developed from human monocytoid cells were selected by ECVAM for prevalidation and validation studies with the intent of comparing their effectiveness for replacing the RPT and thereby eliminating the use of rabbits for pyrogen testing. The results of these studies have been published (Hoffmann et al. 2005a; Schindler et al. 2006). The five tests selected were:

- The Human WB/IL-1β *In Vitro* Pyrogen Test[2]

- The Human WB/IL-1β *In Vitro* Pyrogen Test: Application of Cryopreserved (Cryo) Human WB
- The Human WB/IL-6 *In Vitro* Pyrogen Test
- The Human PBMC/IL-6 *In Vitro* Pyrogen Test[2]
- The MM6/IL-6 *In Vitro* Pyrogen Test

1.1.2 Peer Reviews of In Vitro Pyrogen Test Method Validation Studies

The ECVAM-sponsored validation studies of each of these *in vitro* test methods have been the subject of a recent formal peer review convened by the ECVAM Scientific Advisory Committee (ESAC). Two members of the ESAC served as co-chairpersons for the review Panel, which consisted of five additional U.S. and European reviewers. These reviewers assessed the ability of each test method to serve as a complete replacement for the RPT. Based on this review, the ESAC declared that, "these tests have been scientifically validated for the detection of pyrogenicity mediated by Gram-negative endotoxins, and quantification of this pyrogen, in materials currently evaluated and characterized by RPTs." Although the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) requested the ESAC peer review report, this document is not currently permitted by ECVAM to be publicly disseminated.

This BRD was prepared for an ICCVAM independent peer review panel (Panel) to evaluate these *in vitro* pyrogen test methods and to consider the ICCVAM draft recommendations for each *in vitro* test method. Because individual BRDs for each test method were provided by ECVAM, the ICCVAM BRD provides information that was common to all five *in vitro* test methods and references the appropriate sections of the ECVAM BRDs for specifics related to individual test methods. The recommendations of the ICCVAM Panel, combined with the information and analyses presented in the ICCVAM and ECVAM BRDs and any comments by the public or the Scientific Advisory Committee on Alternative Toxicological Methods were considered by ICCVAM prior to making its final recommendations on the usefulness and limitations of each test method, the proposed standardized test method protocols, performance standards, and any additional studies considered necessary to further develop or characterize any or all of these *in vitro* test methods.

1.2 Regulatory Rationale and Applicability

1.2.1 Current Regulatory Testing Requirements and ICCVAM Prioritization Criteria

This section reviews and summarizes the extent to which the five ICCVAM prioritization criteria (ICCVAM 2003) apply to the *in vitro* pyrogen test methods under consideration.

Criteria 1. The extents to which the proposed test methods are (a) applicable to regulatory testing needs and (b) applicable to multiple agencies/programs.

---

[2]As indicated in the ECVAM BRDs for the WB/IL-1 and PBMC/IL-6 test methods, catch-up validation studies were conducted to evaluate the performance of the WB/IL-1 test method when using 96-well plates, and the PBMC/IL-6 test method when using cryopreserved PBMCs. The plating procedure (WB/IL-1) and the cryopreservation procedure (PBMC/IL-6) are the only differences in the test method protocols (see **Appendix A**). These modifications were not submitted by ECVAM as separate test methods, and are therefore not being considered as separate test methods in this BRD. However, where relevant, comparative information is provided (e.g., see **Table 2-1** and **Sections 6.1** and **7.2**).

Pyrogenicity testing is primarily used by regulatory authorities for end-product release of human and animal parenteral drugs, biological products, and medical devices. The results from these assays are used to limit, to an acceptable level, the risks of febrile reaction in the patient exposed to the product of concern by injection and/or implantation. As detailed in **Table 1-1**, the U.S. Food and Drug Administration (FDA) is the principal U.S. regulatory agency that requires pyrogenicity testing, with different Centers within the FDA regulating the affected products. The Center for Biologics Evaluation and Research, the Center for Drug Evaluation and Research, the Center for Devices and Radiological Health, and the Center for Veterinary Medicine require that human injectable drugs (including biological products), animal injectable drugs, and medical devices be tested for the presence of pyrogenic substances. The current U.S. legislation requiring the use of pyrogenicity testing is set forth in the Federal Food, Drug, and Cosmetic Act (U.S. Code [U.S.C.], Title 21, Chapter 9). In addition, the U.S. Pharmacopeia (USP) maintains sterility requirements for pharmaceuticals that include pyrogenicity testing. **Table 1-1** also shows the statutory protocol requirements used by each FDA Center, along with the comparable enabling legislation and statutory protocol requirements of the EU member nations.

Additionally, the prediction model described in the ECVAM BRDs is based on a pyrogen threshold dose of 0.5 EU/mL. While this level of detection would be sufficient for many parenteral drugs and medical devices, the endotoxin limit set by the FDA for intrathecal drugs and devices that contact cerebrospinal fluid is 0.06 EU/ml. In response to an ICCVAM Pyrogenicity Working Group (PWG) request for more data to support the use of these test methods for discriminating an endotoxin threshold dose lower than 0.5 EU/ml, ECVAM provided supplemental data (see question #6 in **Appendix B**).

Table 1-1 Summary of U.S. and European Legislation and Statutory Protocol Requirements for Pyrogenicity Testing

Agency	Regulated Products	Legislation	Statutory Protocol Requirements	Non-Governmental Standards
United States				
FDA-CBER	Biological products	Federal Food, Drug, and Cosmetic Act (U.S.C. Title 21, Chapter 9)	21 CFR 610.13 (FDA 2005)	USP30 NF25<85> (USP 2007a) USP30 NF25<151> (USP 2007b) ISO 10993-11 (ISO 2006)
FDA-CDER	Human parenteral pharmaceuticals			
FDA-CDRH	Medical devices			
FDA-CVM	Veterinary pharmaceuticals			
Europe				
EDQM	Human/veterinary parenteral pharmaceuticals, biological products, medical devices	Council Regulation (EEC) 230/9/93 Council Directive 93/39/EEC Council Directive 93/40/EEC	EP5.0 2.6.8 (EP 2005a) EP5.0 2.6.14 (EP 2005b)	ISO 10993-11 (ISO 2006)
EMEA				
Regulatory Authorities for Individual EU Countries				

Abbreviations: CBER = Center for Biologics Evaluation and Research; CDER = Center for Drug Evaluation and Research; CDRH = Center for Devices and Radiological Health; CFR = U.S. Code of Federal Regulations; CVM = Center for Veterinary Medicine; EDQM = European Directorate for the Quality of Medicines; EEC = European Economic Community; EMEA = European Medicines Agency; EP = European Pharmacopeia; EU = European Union; FDA = U.S. Food and Drug Administration; ISO = International Standards Organization; NF = National Formulary; U.S.C. = United States Code; USP = U.S. Pharmacopeia

Criteria 2. Warranted, based on the extent of expected use or application and impact on human, animal, or ecological health.

The proposed test methods are intended to replace a method that is used extensively in pharmaceutical, biological product, and medical device development and registration (i.e., the RPT).

Criteria 3: The potential for the proposed test methods, compared to current test methods accepted by regulatory agencies, to (a) refine animal use (decrease or eliminate pain and distress), (b) reduce animal use, or (c) replace animal use.[3]

The two most common pyrogen tests presently used (i.e., RPT and BET) require the use of animals. The RPT is performed in rabbits that can be maintained and reused (under certain circumstances) for multiple tests. According to USP30 NF25<151> (USP 2007b), rabbits may not be reused more than once every 48 hours (hr) after a negative test, not less than two

[3]*Refinement alternative* is defined as a new or revised test method that refines procedures to lessen or eliminate pain or distress to animals, or enhances animal well-being. *Reduction alternative* is defined as a new or revised test method that reduces the number of animals required. *Replacement alternative* is defined as a new or revised test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate) (ICCVAM 1997).

weeks following either (1) a maximum rise of 0.6°C or more, or (2) an animal is included in a test with a substance that is classified as pyrogenic. The BET is performed using hemolymph (the equivalent of blood, which requires drawing approximately 20% of the animal's total blood volume) obtained from *Limulus polyphemus* (horseshoe crabs). Although the donor horseshoe crabs are returned to the wild, some mortality (up to 15%) is associated with the procedure (Walls et al. 2002).

The need for horseshoe crab hemolymph has potentially been reduced with the development of recombinant Factor C, the endotoxin sensitive protein that initiates clotting in the traditional BET. This commercial product, which was originally cloned from the horseshoe crab (Ding et al. 1997; Ding and Ho 1998, 2001), is currently being compared to the BET for submission and inclusion in the USP.

It should also be noted that the FDA has accepted data from the PBMC test developed by Novartis and Baxter Healthcare, which in conjunction with RPT and BET results, were used to support the safety testing of a specific single drug product (New Drug Application Number 16-267/S-037).

Criteria 4: The potential for the proposed test methods to provide improved prediction of adverse health or environmental effects, compared to current test methods accepted by regulatory agencies.

Sufficient data are presented to allow an assessment of the performance of the proposed test methods relative to the RPT (see **Section 6.0**). Because these methods are conducted using cells of human origin, it is postulated that they may reflect the human physiological response better than the currently employed, non-human based methods (i.e., RPT and BET).

These *in vitro* test methods have also been found to be useful for detecting test substances that, in the absence of endotoxin contamination, have evoked an adverse response in patients (Marth and Kleinhappl 2002; Martis et al. 2005). For example, numerous cases of aseptic peritionitis in dialysis patients that were not febrile have been attributed to peptidoglycan contamination of the dialysate (Martis et al. 2005). This dialysate solution met all European and USP standards prior to product release, but the PBMC/IL-6 test method detected increased levels of IL-6 when the dialysate was tested following product recall (Martis et al. 2005). In Marth and Kleinhappl (2002), a case study of a vaccine that was approved for release by the Austrian health authorities, but later produced a fever response in humans, has been described. When this vaccine was subsequently tested in the WB/IL-1β test method, it produced a positive result (Marth and Kleinhappl 2002).

Criteria 5: The extent to which the test method provides other advantages (e.g., reduced cost and time to perform) compared to current methods.

As outlined in **Table 11-1**, cost estimates obtained from various contract laboratories that perform the RPT and from the only contract laboratory known to perform an *in vitro* pyrogen test indicate that the *in vitro* test methods are considerably more cost effective to perform than the RPT. With respect to time considerations, the *in vitro* test methods require two half-days (i.e., one before and one after the overnight incubation) to complete, provided that cryopreserved blood is available and that interference testing is not required. The RPT can be performed within one day. However, before using a rabbit for the first time in a RPT, it must

be conditioned by a sham test that includes all steps of pyrogenicity testing except for injection, according to USP30 NF25<151> (USP 2007b).

1.2.2 Intended Uses of the Proposed In Vitro Pyrogen Test Methods

The proposed test methods are intended as an end-product release test for the identification of Gram-negative endotoxin in human and animal parenteral drugs, biological products, and medical devices. The results from pyrogen testing are used to limit, to an acceptable level, the risks of febrile reaction to the injection and/or implantation of the product of concern.

1.2.3 Similarities and Differences in the Endpoints Measured by the Proposed Test Methods and the In Vivo Reference Test Method

The endpoint measured in the *in vitro* pyrogen test methods is release of proinflammatory cytokines, either IL-1β or IL-6, in response to a test substance challenge, depending on the specific cell type employed. As described in **Section 1.1.1**, the pathogenesis of fever is induced by bacterial products that stimulate the production of IL-1α, IL-1β, and TNF-α, which leads to secondary synthesis of IL-6 and subsequent induction of prostaglandin synthesis (Netea et al. 2000). Direct injection of either IL-1 or IL-6 in several species causes fever, but much higher concentrations of IL-6 are needed. For example, in the rabbit, up to 100-fold more IL-6 is needed to produce a fever compared to IL-1 (Dinarello 2004).

The RPT involves measuring the rise in body temperature evoked in rabbits by the intravenous (i.v.) injection of a test solution. Although there is no direct association between the endpoints measured in the *in vitro* test methods and the RPT, fever is mediated by proinflammatory cytokines and therefore, it is reasonable to postulate that the cytokine-inducing potential of a pyrogen should correlate with its pyrogenic potential (Nakagawa et al. 2002). Moreover, Nakagawa et al. (2002) evaluated the utility of *in vitro* pyrogen test methods for detecting and quantifying various pyrogens. For example, the authors demonstrated that the responsiveness of human WB cells to pyrogens was very similar to that of a subline of MM6 cells, where endotoxin treatment (1 ng/mL) resulted in the production of IL-6 (~1 ng/mL) and IL-1 (~0.1 ng/mL).

Because the RPT is based solely on a rise in body temperature, no data were found on proinflammatory cytokine levels in rabbits following injection with endotoxin to permit a direct comparison with the *in vitro* test methods.

1.2.4 Use of the Proposed Test Methods in an Overall Strategy of Hazard or Safety Assessment

As detailed in **Table 1-1**, current U.S. and European regulatory requirements exist to test pharmaceutical products, biological products, and medical devices for pyrogenicity. The pyrogen tests that are currently acceptable to regulatory authorities require the use of rabbits or horseshoe crab hemolymph. According to ECVAM, the *in vitro* test methods are intended to replace the RPT for the identification of pyrogens where: (a) the test material is incompatible with the BET or (b) the test material contains a non-endotoxin mediated pyrogen. However, as detailed in **Section 3.0**, only Gram-negative endotoxin was included in the validation study. Therefore, other types of pyrogens have not been adequately validated

(refer to **Section 1.1.2**)[4]. The extent to which the RPT is performed only for detecting the presence of endotoxin is not clear.

1.3 Scientific Basis for the *In Vitro* Pyrogen Test Methods

1.3.1 Purpose and Mechanistic Basis of the In Vitro Pyrogen Test Methods

The proposed methods are intended to detect pyrogens in parenteral pharmaceuticals, medical devices, and human biological products. These methods provide an *in vitro* model of the initiation of the human fever response by measuring proinflammatory cytokine release (i.e., IL-1β or IL-6) from human monocytes/monocytoid cells exposed to pyrogens. These proinflammatory cytokines are associated with the initiation of the *in vivo* fever response.

1.3.2 Similarities and Differences of Modes of Action Between the In Vitro Pyrogen Test Methods and the Fever Response in Humans and/or Rabbits

As detailed in **Section 1.2.3**, each of the five proposed *in vitro* test methods measure proinflammatory cytokine release (i.e., IL-1β or IL-6) from human monocytoid cells as an indicator of the presence of a pyrogenic substance. By comparison, the RPT measures a change in body temperature in rabbits over a specified time period following an i.v. injection of a test substance. Although the relative sensitivities of each species to Gram-negative endotoxins vary, the responses of humans, horseshoe crabs (via hemolymph gelatin), and rabbits to these pyrogens have been studied extensively, and test methods based on blood products or blood cells from each of these species appear to be capable of responding to pyrogens (Greisman and Hornick 1969; Cooper et al. 1971; Brunson and Watson 1974; Hoffman et al. 2005a). Several studies directly comparing the *in vitro* pyrogen test methods with either the RPT and/or BET are summarized in **Section 9.1**, Moseby et al. (2000), and in the ECVAM response to ICCVAM PWG questions (see question #1 in **Appendix B**).

The recent discovery and characterization of the TLR family, which recognizes a diverse range of molecules such as lipids, proteins, and nucleic acids derived from pathogens, has led to an enhanced understanding of the signaling pathways activated by endotoxin. More specifically, TLR-4 has been identified as the receptor directly utilized by endotoxin to elicit an immune response. Upon recognition of endotoxin, TLR-4 initiates a rapid and complex signaling cascade, which activates transcription factors (i.e., NF-κB, AP-1, and interferon regulatory factors) to produce proinflammatory cytokines and other immune modulators, thereby leading to a protective immune response (Ishii et al. 2005; Ishii and Akira 2006). It is important to recognize that this TLR-dependent production of proinflammatory cytokines is distinct from the endotoxin-induced synthesis of IL-1 that then converges on the same signaling pathway via the IL-1 receptor (Conti et al. 2004). In addition to endotoxin, TLR-4 recognizes numerous other microbial components such as respiratory syncytial virus proteins and anthrolysin O (Ishii et al. 2005, Ishii and Akira 2006). When proinflammatory cytokine mRNA levels (i.e., IL-1, IL-6, and TNF-α) were compared in response to various TLR-4 agonists, endotoxin induced the highest level of expression (Park et al. 2004).

[4]Additional information on testing of other types of pyrogens was also provided by ECVAM and is included in **Appendix B**.

It has been recognized for many years that humans are responsive to relatively low doses of endotoxin, whereas rodents require much higher doses to elicit a response. In recent years, these species differences have been attributed, in part, to structural differences in TLR-4. For this reason, caution should be used when extrapolating findings from other mammals to humans with respect to endotoxin and TLR-4 signaling (Stoll et al. 2006). Furthermore, TLR-4 mutations have been identified in mice and humans, and it is likely that such defects are associated with altered gene expression and increased susceptibility to infection (Norata et al. 2005; van Deventer 2000; von Aulock et al. 2003).

1.3.3 Range of Substances Amenable to the In Vitro Pyrogen Test Methods and Limits of These Methods

The proposed methods are intended for the identification of pyrogenic substances in parenteral pharmaceuticals, biological products, and medical devices. Because they are based on cultured human monocytes/monocytoid cells, they are considered capable of detecting both Gram-negative endotoxin and non-endotoxin-based pyrogens. While **Section 9.0**, Moesby et al. (2005), and the ECVAM response to ICCVAM PWG questions (see question #2 in **Appendix B**) provide a number of published studies demonstrating that the *in vitro* pyrogen test methods are able to detect non-endotoxin pyrogens, the ECVAM validation studies focused specifically on Gram-negative endotoxin due to the unavailability of standardized, non-endotoxin pyrogens (see **Section 3.0**).

Because these test methods measure the release of proinflammatory cytokines, drugs that are cytotoxic to blood cells or that induce a substantial proinflammatory response (e.g., IL-1 receptor antagonists, interferon [IFN]- γ , and rheumatic factors) are not amenable to testing by these methods (Hartung et al. 2001; Ishii et al. 2005; Ishii and Akira 2006). As described in **Section 2.0**, each test method includes an interference test to identify problematic test samples.

1.4 Validation of the *In Vitro* Pyrogen Test Methods

The ICCVAM Authorization Act of 2000 mandates that "[each] Federal Agency … shall ensure that any new or revised … test method … is determined to be valid for its proposed use prior to requiring, recommending, or encouraging [its use]." Validation is the process by which the reliability and relevance of an assay for a specific purpose are established (ICCVAM 1997). Relevance is defined as the extent to which an assay will correctly predict or measure the biological effect of interest (ICCVAM 1997). For the *in vitro* pyrogen test methods described in this ICCVAM BRD, relevance is restricted to how well the assays detect the presence of Gram-negative endotoxin. Reliability is defined as the reproducibility of a test method within and among laboratories and should be based on performance with a diverse set of substances that are representative of the types of chemical and product classes that are to be tested and the range of responses that needs to be identified. The validation process is designed to provide data and information that will allow ICCVAM to make recommendations on the applicability of a test method and U.S. Federal agencies to consider those recommendations in light of their regulatory mandates.

The first stage in the evaluation of a new test procedure is the preparation of a BRD that presents and evaluates the relevant data and information about the test method, including its

mechanistic basis, proposed uses, reliability, and performance characteristics (ICCVAM 1997). This ICCVAM BRD summarizes the available information on each of the five *in vitro* pyrogen test methods listed in **Section 1.1.1**.

Where adequate data are available, the qualitative and quantitative performance of the proposed alternative test method is evaluated, and its reliability is compared with the reliability of the currently accepted test method. This ICCVAM BRD will aid in identifying essential test method components that should be considered during the identification of a standardized protocol for use of the test method.

1.5 Search Strategies and Selection of Citations for the ICCVAM *In Vitro* Pyrogen Test Methods BRD

NICEATM conducted an online literature search for relevant information on the five *in vitro* pyrogen test methods using multiple internet databases (i.e., PubMed, SCOPUS, TOXLINE, Web of Science). Specifically, records were sought using various combinations of the terms: *in vitro*, WB, WB cells, PBMC, Mono Mac 6, MM6, endotoxin, lipopolysaccharide (LPS), pyrogen, LAL, BET, IL-1, and IL-6. This search was conducted to supplement and update the list of peer-reviewed publications related to *in vitro* pyrogen testing that was provided in the ECVAM BRDs. U.S., EU, and Japanese pyrogenicity test guidelines were obtained from relevant regulatory agencies via the internet or through direct requests. The resulting database of 370 references confirmed that the lists of references included in the ECVAM BRDs were complete and up-to-date.

2.0 *In Vitro* Pyrogen Test Method Protocol Components

2.1 Overview of How the *In Vitro* Pyrogen Test Methods Are Conducted

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

- Interference testing is performed to verify that a test substance does not interfere with either the cell system used or with the specific cytokine-specific enzyme-linked immunosorbent assay (ELISA).
- The test substance is mixed with a suspension of human-derived blood cells.
- The concentration of the specific proinflammatory cytokine (e.g., IL-1β, IL-6) is measured using an ELISA, and is compared to the response curve of an endotoxin standard.
- An internationally accepted endotoxin standard (World Health Organization-LPS [WHO-LPS] 94/580 *Escherichia coli* [*E. coli*] O113:H10:K-), or an endotoxin standard that has been calibrated against this standard, is used to generate the standard response curve for the assay. The endotoxin activity of a test substance is calculated by comparing the induced cytokine release with that induced by the endotoxin standard.
- A product "passes" (i.e., is considered negative for endotoxin pyrogen activity) if the cytokine response to the test substance is less than that induced by 0.5 endotoxin units/mL (EU/mL).

2.2 Description and Rationale for the Test Method Components for Proposed Standardized Protocols

The standard operating procedures for each test method assessed in the ECVAM validation studies are provided as Appendix A of each ECVAM BRD. As indicated in **Section 2.1**, there are essential principles of each protocol that are common among the five methods reviewed. These include:

- Isolating and/or culturing human monocytoid cells (either included in WB, separated as a fraction [i.e., PBMCs], or as cell line [i.e., MM6])
- Performing interference testing with each substance
- Treating the cells in suspension with a test substance
- Collecting cytokine release data
- Evaluating the data in relation to the proposed prediction model

Table 2-1 provides a comprehensive comparison of the similarities and differences among the protocols for the five test methods. No rationale was provided for the use of WB in the various test methods; however, Poole et al. (2003) summarized several studies, which indicated that the monocytes present in diluted WB respond to pyrogen/endotoxin by releasing pyrogenic cytokines. The use of the MM6 cell line was justified based on mechanistic considerations and its response to endotoxins.

[This Page Intentionally Left Blank]

Table 2-1 *In Vitro* Pyrogen Test Method Components

	Test Method Component	WB/IL-1β[1]	Cryo WB/IL-1β	WB/IL-6	PBMC/IL-6[1]	MM6/IL-6
	Source of cells	Human WB	Human Cryo WB	Human WB	Human WB	MM6 cell bank (original cell line maintained by Prof. H. Ziegler-Heitbrock, U. Munich)
	Laboratory equipment	• CO_2 cell culture incubator (37°C, 5% CO_2, humidified) • Centrifuge • Consumables as specified in SOP (e.g., heparinized blood tubes for WB methods, 96-well plates or culture tubes, centrifuge tubes, microfuge tubes, pyrogen-free plastic-ware where specified, serological pipets, pyrogen-free hypodermic needles) • Data analysis software • Hemocytometer (e.g., PBMC and MM6 assays) • Laminar Flow Hood (Class II) • Liquid nitrogen, CO_2 freezers, or programmable freezers for cryopreservation methods (Cryo WB/IL-1β or Cryo PBMC) • Microscope, inverted (optional except for PBMC and MM6 assays) • Microtiter Plate Reader (450 nm with 600-690 nm reference filter for IL-1β or 500-590 nm reference filter for IL-6 measurements) • pH meter • Pipettors (8 to 12 multi-channels; 2 to 2000 µL adjustables; pyrogen-free tips (except for ELISA) • Vortex mixer • Water bath				
	Culture medium	None - WB is diluted with PFS	RPMI Complete Medium • RPMI 1640 (part of the Endosafe Kit for cryoblood) • no specific additives needed	None - WB is diluted with PFS	RPMI Complete Medium • RPMI 1640 • HSA • L-Glutamine (2 mM) • Penicillin/streptomycin	RPMI Medium[2] • RPMI 1640 medium • Bovine insulin (0.23 IU/mL) • HEPES (20 mM) • HIFCS (10% or 2%) • L-glutamine (2mM) • MEM non-essential amino acid solution (0.1 mM) • Oxaloacetic acid (1 mM) • Sodium pyruvate (1 mM)
	Other reagents	• Endotoxin standard • PFS • PFW • Validated IL-1β ELISA kit	• DMSO • Endotoxin standard • PFS • PFW • Validated IL-1β ELISA kit	• Endotoxin standard • PFS • PFW • Validated IL-6 ELISA kit	• Endotoxin standard • PFS • PFW • Trypan blue • Validated IL-6 ELISA kit	• DMSO • Endotoxin standard • PFS • PFW • Trypan blue • Validated IL-6 ELISA kit
Dose selection procedures		Interference testing performed to determine the lowest dilution of the test product necessary to achieve an acceptable endotoxin spike recovery (i.e., 50% to 200% recovery)[3]				
Endpoints measured		IL-1β release via ELISA		IL-6 release via ELISA		

	Test Method Component	WB/IL-1β[1]	Cryo WB/IL-1β	WB/IL-6	PBMC/IL-6[1]	MM6/IL-6
Exposure of the test substance	**Pre-test preparation of cells**	Collect WB, heparinize, and use within 4 hr Plate Method: same collection procedure	• Collect WB, heparinize, and cryopreserve according to the Konstanz or PEI method • Prior to testing, thaw WB at 37°C for 15 min	Collect WB, heparinize, and use within 4 hr	• Collect WB and isolate PBMCs by centrifugation • Resuspend PBMCs in RPMI-C ($1x10^6$ cells/mL) (use PBMCs within 4 hr of initial WB collection)	• Incubate MM6 cells ($4x10^5$ cells/mL media) for 24 hr • Resuspend cells ($2.5x10^6$ cells/mL)[3] prior to testing
	Application of the test substance	Tube method: In a microfuge tube mix 1000µL PFS+100µL sample+100µL WB Plate method: In a 96-well plate mix 200µL PFS+20µL sample+20µL WB	Konstanz method: In a 96-well plate mix 200µL RPMI+20µL sample+20µL WB PEI Method: In a 96-well plate mix 180µL RPMI + 20µL sample+40µL WB	In a 96-well plate: Mix 50 µL standards/samples+100 µL PFS+50 µL WB	In a 96-well plate: Mix 50 µL standards/samples+100 µL RPMI-C+100 µL PBMCs	In a 96-well plate: Mix 50 µL standards/samples+100 µL RPMI-C+100 µL cells in suspension
	Duration of exposure	10-24 hr		16-24 hr		
	Material used for ELISA	Tube method: centrifuge 2 min @ 10,000 x *g*-test supernatant Plate method: mix each well be pipetting and test resuspended mixture	WB/RPMI/sample mixture	WB/saline/sample mixture	Cell supernatant	Cell supernatant
Known limits of use		Intended for parenteral pharmaceuticals, biological products, and medical devices that have been qualified through interference testing				
Nature of the response assessed		Pyrogenic substances induce the release of proinflammatory cytokines (e.g., IL-1β) from monocytoid cells present in human WB		Pyrogenic substances induce the release of proinflammatory cytokines (e.g., IL-6) from monocytoid cells in WB, PBMC, or immortalized MM6 cells		
Appropriate controls	**Positive control (PC)**	0.5 EU/mL WHO-LPS 94/580 [*E. coli* 0113:h10:K-][5]				
	Negative control (NC)	PFS				
	Positive product control (PPC)	Test substance spiked with endotoxin (0.5 EU/mL or a concentration in middle of standard endotoxin curve)				
	Negative product control (NPC)	Test substance spiked with PFS				
Assay acceptability criteria		• PC OD 1.6-fold>NC OD • PPC OD 1.6-fold>NPC OD • PPC OD should be within 50% to 200% of the PC OD • NC OD≤0.100	• PC OD 1.6-fold>NC OD • PPC OD 1.6-fold>NPC OD • PPC OD should be within 50% to 200% of the PC OD • NC OD≤0.100	• PPC OD should be within 50% to 200% of the PC OD • NC OD<200 pg/mL IL-6 standard	• PPC OD should be within 50% to 200% of the PC OD • 1 EU/mL standard OD>1000 pg/mL IL-6 standard • NC OD<0.15 and NC OD<500 pg/mL IL-6 standard	• PC OD±20% of the expected value (i.e., 0.5 EU/mL) • PPC OD should be within 50% to 200% of the PC OD • NC OD<0.200

	Test Method Component	WB/IL-1β[1]	Cryo WB/IL-1β	WB/IL-6	PBMC/IL-6[1]	MM6/IL-6
Nature of data to be collected and methods used for data collection		• The endotoxin content of a test substance is calculated by comparing the induced IL-1β release with that induced by the endotoxin standard curve concentrations		• The endotoxin content of a test substance is calculated by comparing the induced IL-6 release with that induced by the endotoxin standard curve concentrations		
Type of media in which data are stored		Electronic files				
Exclusion criteria		Mean±SD of the OD for each test substance/standard				
Decision criteria for pyrogenicity		OD TS > OD 0.5 EU/mL EC		EC TS > ELC TS	EC TS > ELC TS[6]	EC TS > ELC TS

Abbreviations: Cryo = Cryopreserved; DMSO = Dimethylsulfoxide; EC = Endotoxin concentration; ELC = Endotoxin limit concentration; EU/mL = Endotoxin units/mL; ELISA = Enzyme-linked immunosorbent assay; HIFCS = Heat-inactivated fetal calf serum; HSA = Human serum albumin; IL = Interleukin; IU = International units; LPS = Lipopolysaccharide; MEM = Minimum essential medium; MM6 = Mono Mac 6; NC = Negative control; NPC = Negative product control; OD = Optical density; PBMC = Peripheral blood mononuclear cells; PC = Positive control; PEI = Paul-Ehrlich-Institut; PFS = Pyrogen free saline; PFW = Pyrogen free water; PPC = Positive product control; SD = Standard deviation; SOP = Standard operating procedure; TS = Test substance; WB = Whole blood; WHO = World Health Organization; x *g* = times gravity

[1]As described in **Section 1.1.1**, a catch-up validation studies were also conducted to evaluate the performance of the WB/IL-1β test method using 96-well plates, and the PBMC/IL-6 test method when using cryopreserved PBMCs. The plating procedure (WB/IL-1β) and the cryopreservation procedure (PBMC/IL-6) are the only differences in the test method protocols (see **Appendix A**).

[2]Medium should be qualified for testing by a valid bacterial endotoxin test (i.e., USP30 NF25<85>) indicating that the endotoxin contamination is <0.06 IU/mL); fetal bovine serum concentration for MM6 cells varies based on whether it is for maintenance/propagation (10%) or assay (2%) conditions.

[3]Dilution of the test material should not exceed the maximum valid dilution (MVD), where MVD = (endotoxin limit concentration)/(detection limit of the assay)

[4]Cell numbers represent viable cells based on trypan blue exclusion

[5]Or another endotoxin calibrated against this standard

[6]Includes a sequential decision strategy in which 3 to 4 donors are tested per substance. 1) If all donors show negative - product is non-pyrogenic; 2) If ≥2 donors show a positive - product is pyrogenic; 3) If only one donor shows a positive, an additional 3 to 4 donors are tested and if no more than one donor is positive (out of 6 to 8 donors) - product is non-pyrogenic; otherwise, product is pyrogenic**.**

[This Page Intentionally Left Blank]

2.2.1 Methods Used to Analyze the Data, Including Methods to Analyze for Interference with the Assay

Once a substance has been tested in the requisite number of donor samples (see **Section 2.2.2**), the resulting sample test medium (as indicated in **Table 2-1**) is assayed in quadruplicate in the relevant cytokine ELISA. Outliers are identified using the nonparametric Dixon's test ($p > 0.05$) (Dixon 1950; Barnett and Lewis 1984), the Grubbs' test (Grubbs 1969) for normally distributed samples, or other statistically acceptable methods (Martin and Roberts 2006) and are excluded from the calculations of endotoxin content (see also **Section 5.3** and **Appendix C**). Endotoxin standard curves are included in each assay, from which the endotoxin content of each replicate is estimated using a 4-parameter logistic model.

As indicated in **Table 2-1**, mean optical density (OD) readings are calculated for the positive and negative control samples, as well as for the relevant positive and negative product controls. The acceptable range of the positive product control (50% to 200% of the positive control response) defines the threshold for interference with the test system. If the positive product control response falls outside of this range, the samples are then assayed at the lowest dilution that does not cause interference.

2.2.2 Decision Criteria and the Basis for the Prediction Model Used to Identify a Pyrogenic Substance

As described in **Section 4.2**, historical RPT data were used to establish a threshold pyrogen dose (i.e., the endotoxin dose at which fever was induced in 50% of the rabbits), which was determined to be 5 EU/kg. Based on the largest allowable volume for injection in rabbits (10 mL/kg), the limit of detection that the *in vitro* pyrogen tests must meet was defined as 0.5 EU/mL. Accordingly, the prediction model for each test method was established based on this limit of detection (i.e., a substance is considered pyrogenic if the mean response is greater than or equal to the 0.5 EU/mL standard).

For three of the test methods, results from multiple donors (Cryo WB/IL-1β [n=5], WB/IL-6 [n=3], and PBMC/IL-6 [n=3 to 4]) are required to determine the potential pyrogenicity of a test substance. In contrast, a single donor sample is used for the WB/IL-1β test method, as is a single cell sample for the MM6/IL-6 test method. As outlined in **Table 2-2**, unlike the Cryo WB/IL-1β test method, the WB/IL-6 and PBMC/IL-6 test methods employ a decision strategy that takes into account the individual responses of each donor sample.

Table 2-2 Prediction Model Used for *In Vitro* Pyrogen Test Methods

Test Method	No. Donors	No. Positive	No. Negative	Decision
PBMC/IL-6	4[1]	4	0	Pyrogenic
		3	1	Pyrogenic
		2	2	Pyrogenic
		1	3	Non-pyrogenic
		0	4	Non-pyrogenic
	3[1]	3	0	Pyrogenic
		2	1	Pyrogenic
		1	2	Non-pyrogenic
		0	3	Non-pyrogenic
WB/IL-6	3	3	3	Pyrogenic
		2	1	Pyrogenic
		1	2	Non-pyrogenic
		0	3	Non-pyrogenic
Cryo WB/IL-1β	5 (pooled)[2]	1	0	Pyrogenic
		0	1	Non-pyrogenic
WB/IL-1β	1	1	0	Pyrogenic
		0	1	Non-pyrogenic
MM6/IL-6	NA[3]	1	0	Pyrogenic
		0	1	Non-pyrogenic

Abbreviations: Cryo = Cryopreserved; IL = Interleukin; MM6 = Mono Mac 6; NA = Not applicable; PBMC = Peripheral blood mononuclear cells; WB = Whole blood

[1]Samples are collected from four donors for the PBMC/IL-6 test method. One donor sample may be excluded based on quality criteria, in which case the prediction model may be applied to results from three donors.

[2]Samples are collected from five donors for the Cryo WB/IL-1β test method and pooled prior to cryopreservation.

[3]Not applicable, because source material is obtained from an immortalized cell line.

2.2.3 Information and Data to be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission

The test report should include the following information, if relevant to the conduct of the study:

Test Substances and Control Substances

- Name and type (e.g., pharmaceutical, biological product, medical device eluate, etc.) of test product
- Purity and composition of the test substance or preparation
- Physicochemical properties, such as physical state, volatility, pH, stability, chemical class, water solubility, relevant to the conduct of the study
- Quality assurance (QA) data and known biological properties
- Treatment of the test/control substances prior to testing, if applicable (e.g., vortexing, sonication, warming; solvent used)
- Stability, if known

Justification of the Specific Protocol(s) Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, documentation of the procedure used to ensure their integrity from lot-to-lot and over time

Criteria for an Acceptable Test

- Acceptable concurrent positive control ranges based on historical data from the testing laboratory (which should be included in the report)
- Acceptable negative control data, including historical control ranges from the testing laboratory (which should be included in the report)

Test Conditions

- Cell system used; donor information, if relevant
- Calibration information for the equipment used for measuring cytokine release (e.g., spectrophotometer)
- Details of test procedure used
- Description of modifications of the test procedure made by the testing laboratory for the substance being tested
- Reference to the laboratory's historical data for the cell system and protocol
- Description of data and QA evaluation criteria used

Results

- Tabulation of data from individual test samples

Description of Other Effects Observed

Discussion of the Results

Conclusion

A Good Laboratory Practice (GLP) QA Statement

- This statement addresses all GLP inspections and audits made during the study, and the dates the results were reported to the Study Director. This statement also serves to confirm that the final report reflects the raw data.

Reporting requirements for GLP-compliant studies are provided in the relevant guidelines (e.g., Organisation for Economic Co-operation and Development [OECD] 1998; U.S. Environmental Protection Agency [EPA] 2003a, 2003b; FDA 2003).

2.3 Basis for Selection of the Test Method Systems

One of the difficulties associated with the currently required pyrogen test methods (i.e., BET and RPT) is that both require extrapolation of the response from a non-human system to the human. In contrast, and as discussed in **Section 1.1.1**, all five of these test methods employ human cells in an attempt to mimic the human fever response *in vitro*.

The WB test methods (i.e., Cryo WB/IL-1β, WB/IL-1β, WB/IL-6) offer the convenience of performing the assay directly on a human blood sample, with minimal pretest preparation. The Cryo WB/IL-1β test method was developed to offer the convenience of an increased time interval between the time of blood collection and the time a test is initiated (since the fresh blood methods require testing within four hr of collection), as well as increased standardization through the pooling of five donor samples to produce a larger sample bank of cells to use in the test. The MM6/IL-6 test method provides increased standardization by using an immortalized cell line that may be maintained in the laboratory indefinitely, and transferred among laboratories. Finally, the PBMC/IL-6 test method was developed in an attempt to improve pyrogen detection sensitivity by using the monocyte fraction of WB, which is considered to be the most sensitive human blood cell type to the presence of endotoxin.

Additional information on standardization of the cellular components required for the test methods is presented in the ECVAM response to ICCVAM PWG questions (see question #5 in **Appendix B**).

2.4 Proprietary Components

Data from the test methods that use the IL-6 endpoint (i.e., WB/IL-6, PBMC/IL-6, MM6/IL-6) were obtained using either an in-house IL-6 ELISA developed by Novartis Pharma AG (Basel, Switzerland) or the Central Laboratory for the Blood Transfusion Service (CLB) Human IL-6 ELISA kit (Amsterdam, The Netherlands). In the ECVAM response to ICCVAM PWG questions (**Appendix B**), it was stated that both IL-6 ELISAs use the same monoclonal anti-IL-6 antibody for detection. At the present time, the Novartis IL-6 ELISA is not available for purchase; however, the CLB IL-6 ELISA kit is commercially available. Importantly, other commercially available IL-6 ELISAs may be individually validated and used in these procedures.

The MM6 cell line was generated by Professor Ziegler-Heitbrock at the University of Munich (refer to **Appendix A**, as well as Section 2.4 of the ECVAM MM6/IL-6 BRD). These cells are currently available from The German Collection of Microorganisms and Cell Cultures (DSMZ). However, a legal agreement with Professor Ziegler-Heitbrock stating that the MM6 cell line will be used for research purposes only is required prior to purchasing the cells. At the present time, any organization (e.g., pharmaceutical company) wishing to use the cells for product testing has to negotiate a fee for provision of the cells and a royalty payment per batch of product tested.

According to the U.S. Patent and Trademark Office (USPTO), patents are held for "Test for determining pyrogenic effect of a material" (U.S. 5,891,728, April 6, 1999), and "Pyrogenicity test for use with automated immunoassay systems" (U.S. 6,696,261 B2, February 24, 2004). These patents cover the WB/IL-1β and WB/IL-6 test methods, respectively. In addition, and related to the WB/IL-1β test method, there is a patent application pending for "Test procedure with biological system - Preparations containing deep-frozen blood are used for determining blood response" (USPTO 436518000).

There are several measures in the study validity criteria that may be used to verify the integrity of proprietary components. As outlined in **Table 2-1**, an endotoxin standard curve is established for each assay, which is in turn used to define the endotoxin activity of the test

substances. In addition, positive and negative controls, along with positive and negative product controls, are used for interference testing, and serve as internal controls for each assay.

2.5 Number of Replicates

2.5.1 Number of Donors

There is no rationale provided for the number of donors included for each test method. As described in **Section 2.2.2**, samples from multiple donors are required for three of the test methods. The Cryo WB/IL-1β test method uses pooled blood from five different donors and the WB/IL-6 and PBMC/IL-6 test methods use blood from at least three donors, which are tested individually. In contrast, a single donor sample is used for the WB/IL-1β test method.

2.5.2 Number of Assay Replicates

Once each substance has been tested in the requisite number of donor samples (see **Section 2.2.2**), the resulting sample test medium is assayed in quadruplicate in the relevant cytokine ELISA. As indicated in **Section 2.2.1**, the nonparametric Dixon's test (Dixon 1950; Barnett and Lewis 1984) or Grubbs' test (Grubbs 1969) for normally distributed samples is used to detect outliers among the replicates. Section 2.5 of the ECVAM BRDs states that four replicates were chosen, as it is considered the minimum number for inclusion in Dixon's test.

2.6 Modifications to the Test Method Protocols Based on ECVAM Validation Study Results

In the MM6/IL-6 test method, prevalidation studies demonstrated that pre-incubation of the cells at a defined initial concentration of 2 x 10^7 cells/50 mL RPMI-C for 24 hr greatly improved test method performance. Therefore, this modification was included in the validation study, and subsequently carried forward to the recommended MM6/IL-6 test method protocol.

For the PBMC/IL-6 test method, a single blood donor was initially used as a source of PBMCs. However, the use of PBMCs from four separate donors (assayed individually) was shown to reduce variability, and this modification was carried forward in the recommended PBMC/IL-6 test method protocol.

No modifications were made to the WB/IL-1β, Cryo WB/IL-1β, and WB/IL-6 test method protocols as a result of the prevalidation or validation testing experiences.

2.7 Differences Between Comparable Validated Test Methods with Established Performance Standards

The differences between the *in vitro* pyrogen test methods and the currently accepted pyrogen test methods (i.e., BET and RPT) are described in **Sections 1.2.3** and **1.3.2**.

[This Page Intentionally Left Blank]

3.0 Substances Used for the Validation of *In Vitro* Pyrogen Test Methods

3.1 Rationale for the Substances or Products Selected for Testing

A validation study should evaluate an adequate subset of substances and product types that are to be tested by the proposed test method. In response to a request for additional information, the rationale for the specific test substances selected for inclusion in the validation studies was provided by ECVAM, which included stability of the endotoxin-spike, relevance, availability/feasibility, and cost (see **Appendix C**). Briefly, to maintain the desired concentration of the endotoxin-spike solution over the time period needed for the validation studies, the test substances and the endotoxin-spike solution were provided separately to the test laboratories and mixed prior to testing. As for relevance, only substances intended for i.v. injection were selected. In addition, test substances consisted solely of marketed parenteral pharmaceuticals that were labeled as free from detectable pyrogens such that these data were available for comparison to the validation study results.

3.2 Number of Substances

A total of 13 substances were included in the performance analysis of each of the five *in vitro* test methods. Ten substances, each spiked with four different 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, 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.

3.3 Identification and Description of Substances Tested

As indicated in **Section 3.1**, the test substances selected for use in the validation studies were marketed parenteral pharmaceuticals. **Table 3-1** lists the 10 test substances used to evaluate accuracy, and **Table 3-2** lists the three test substances used to evaluate reproducibility. In response to a request for additional information, ECVAM provided the lot numbers of the substances used in accuracy evaluation for the validation study, which demonstrated that they were identical (**Appendix C**). However, some of the lots tested in the catch-up validation study for the Cryo WB/IL-1β test method were different (i.e., Fenistil and Sostril) because the original lots were no longer available. One test substance (i.e., Orasthin) was no longer available and was replaced with Syntocinon, which contains the same active ingredient.

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

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

3.4 Sample Coding Procedure

According to the ECVAM BRDs (Section 3.4), the 10 test substances and the four spike concentrations used for the evaluation of accuracy were blinded to the testing laboratories. For the reproducibility analyses, although the three spike concentrations were blinded to the participating laboratories, the identities of the three test substances were not.

3.5 Rationale for the Selection of the Recommended Reference Substances

Reference substances are used to assess the accuracy and reliability of a proposed, mechanistically and functionally similar test method and are a representative subset of those used to demonstrate the reliability and accuracy of the validated reference test method (in this case, the RPT). These substances should:

- Represent the range of responses that the validated test method is capable of measuring or predicting
- Have produced consistent results in the validated test method
- Produce responses that reflect the accuracy of the validated test method
- Have well-defined chemical structures and/or compositions
- Be readily available
- Not be associated with excessive hazard or prohibitive disposal costs

For evaluating test method performance, each of the test substances used in the ECVAM validation studies was spiked with a Gram-negative endotoxin standard (WHO-LPS 94/580 [*E. coli* O113:H10:K-]). Two different sources of endotoxin (i.e., *E. coli* EC-5 and *E. coli* EC-6), which were reported to be identical to the WHO standard, were used in the validation studies (Hochstein et al. 1994; Hoffman et al. 2005a). Endotoxin was selected as a "model" pyrogen for inclusion based on its availability in a standardized form and because of the known ability of monocytic cells to respond to endotoxin-based pyrogens. Endotoxin was also used as a positive control and for qualifying the *in vitro* test methods during interference testing. It is also used when performing the BET. As described in **Section 4.0**, the response of the reference test method (i.e., RPT) to endotoxin is well documented. For this reason, the threshold pyrogen dose used for establishing the decision criteria for the *in vitro* test methods was based on historical RPT data. Importantly, no other non-endotoxin-based pyrogenic substances are presently available in a standardized form.

[This Page Intentionally Left Blank]

4.0 *In Vivo* Reference Data for the Assessment of Test Method Accuracy

4.1 Description of the Protocol Used to Generate *In Vivo* Data

4.1.1 The Rabbit Pyrogen Test

The RPT protocols most widely accepted by regulatory agencies are outlined in the USP (USP 2007b), the U.S. Code of Federal Regulations (FDA 2005), the European Pharmacopeia ([EP], EP 2005a), and the Japanese Pharmacopeia ([JP], JP 2001), and are summarized in **Table 4-1**. The RPT involves measuring the temperature increase in rabbits following an i.v. injection (via the ear vein) of a test substance in a dose not to exceed 10 mL/kg injected within a period of not more than 10 min. Initially, three rabbits are injected and the increase (or decrease) in temperature relative to the baseline value is measured at 30-min intervals for up to three hr. The resulting data are used to calculate an overall temperature increase by adding the results from all three animals, which is then used to assign a label of pyrogenic or non-pyrogenic.

[This Page Intentionally Left Blank]

Table 4-1 Test Guidelines for the Rabbit Pyrogen Test

RPT Protocol Component	Reference			
	21 CFR 610.13 (FDA 2005)	**EP5.0 2.6.8 (EP 2005a)**	**JP XIV (JP 2001)**	**USP30 NF25 <151> (USP 2007b)**
Number of rabbits	3 or 8[1]	3, 6, 9, or 12[1]	3 or 8[1]	3 or 8[1]
Rabbit species/strain	Not specified	Not specified	Not specified	Not specified
Exclusion criteria for rabbits during the initial selection of rabbits	• Used in a negative pyrogen test in the preceding 2 days • Used in a pyrogen test in which its temperature rose ≥0.6°C in the preceding 2 weeks	• Weight<1.5 kg • Decreased weight in the preceding week • Used in a negative pyrogen test in the preceding 3 days • Used in a positive pyrogen test in the preceding 3 weeks	• Weight<1.5 kg • Decreased weight in the preceding week • Previously used in a positive pyrogen test • Rabbits from negative pyrogen tests may be reused only when a "as a long a resting period as possible is taken"	• Used in a negative pyrogen test in the preceding 2 days • Used in a pyrogen test in which its temperature rose ≥0.6°C in the preceding 2 weeks
Testing room conditions	20 to 23°C	Within 3°C of the housing quarters (temperature not specified)	20 to 27°C and constant humidity	20 to 23°C
Food/water during test	Food withheld during the test, but water available at all times	Food withheld overnight and until end of the test. Water withheld during the test.	Food withheld beginning several hrs. prior to first temperature recording and until the end of the test.	Food withheld during the test period, but water available at all times
Depth of temperature probe in rectum	Not less than 7.5 cm	Approximately 5 cm	6-9 cm	Not less than 7.5 cm
Preliminary test	≤7 days prior to main test, perform all procedures used for the main test except the injection.	• 1-3 days prior to main test, treat test animals with an injection of warmed (38.5°C) pyrogen-free saline • Record temperature at 90 min prior to injection and every 30 min thereafter up to 3 hr. • Exclude any rabbits with an increase of >0.6°C	Not specified	≤7 days prior to main test, perform all procedures used for the main test except the injection.
Baseline temperature	• Record temperature ≤ 30 min prior to injection • For any group of rabbits, use only if baseline temperatures do not vary>1°C among rabbits • Exclude rabbits with baseline temperature>39.8°C	• Mean of two temperature recordings at 40 min and 10 min prior to injection • Exclude rabbits if variation >0.2°C between measurements noted • Exclude rabbits with initial temperature >39.8°C or <38.0°C	• Record temperature three times at one-hr intervals prior to injection • Assuming no appreciable variability among recordings, use the last recording as the baseline value. • Exclude animals if 2nd and 3rd temperature measurements exceed 39.8°C	• Record temperature ≤30 min prior to injection • For any group of rabbits, use only if baseline temperatures do not vary >1°C among rabbits • Exclude rabbits with baseline >39.8°C
Injection volume	≥3 mL/kg *BUT* ≤10mL/kg	≥0.5 mL/kg *BUT* ≤10mL/kg	10 mL/kg, unless otherwise specified	≤10 mL/kg

Injection time	≤10 min	≤4 min, unless otherwise indicated	Not specified, but injection should occur within 15 min of the third pretest temperature recording	≤10 min
Injection site	Marginal ear vein	Marginal ear vein	Marginal ear vein	Marginal ear vein
Pre-warming of test material	37°C±2°C	38.5°C	37°C	37°C±2°C
Temperature recording intervals after injection	30 min intervals for 1 to 3 hr	≤30 min intervals for 3 hr	1 hr intervals for 3 hr	30 min intervals for 1 to 3 hr

Abbreviations: CFR = U.S. Code of Federal Regulations; EP = European Pharmacopeia; FDA = U.S. Food and Drug Administration; JP = Japanese Pharmacopeia; RPT = Rabbit pyrogen test; USP = United States Pharmacopeia

[1]Each test is initially conducted with three animals and additional animals are tested to resolve equivocal results in the first three animals

4.1.2 Current In Vivo Pyrogen Test Method Protocols

As indicated in **Table 4-1**, U.S. and international regulatory agencies have tailored the RPT protocol to suit their specific needs and goals in protecting human health. The current test method protocols (i.e., FDA 2005; EP 2005a; JP 2001; USP 2007b) recommend using healthy, adult rabbits with no specific breed/strain requirements. Rabbits are to be adequately acclimated to their surroundings and housed in an environment free from excessive external stimuli. Each rabbit is conditioned prior to the test with a sham test that includes all of the procedural steps except the injection (see also **Section 1.2**). Reuse of test rabbits is permitted only after an appropriate withdrawal period has been completed (see also **Section 1.2**).

The test is conducted in a room that is designated solely for pyrogen testing, in which the temperature is within 3°C of the uniform temperature of the housing room (i.e., 20°C±3°C). Food is withheld during the test, but access to water is continuous. The baseline temperature, which is used to calculate the increase in temperature during the test, is measured 30-40 min prior to injection of the test substance. In each group of rabbits tested, the variation in baseline temperature among the rabbits should not vary more than 1°C, and rabbits with an initial temperature greater than 39.8°C are excluded from testing.

The test substance is pre-warmed to approximately 37°C and injected (≤10 mL/kg) into the marginal ear vein, completing each injection within 10 min. The rectal temperature is recorded at 30-min intervals for up to three hr after the injection. The decision criteria outlined in **Table 4-2** are then used to determine a pyrogenic response. As shown in **Table 4-2**, the decision criteria by which labels of pyrogenic or non-pyrogenic are assigned vary among the USP, FDA, EP, and JP test guidelines.

Table 4-2 Decision Criteria for Determining a Pyrogenic Response in the Rabbit Pyrogen Test

RPT Protocol	No. Rabbits	Product passes if:	Product fails if:
USP30 NF25<151> (USP 2007b)	3	0/3 rabbits show an increase of ≥0.5°C	NA[1]
	5[1]	≤3/8 rabbits show an increase of ≥0.5°C *AND* the summed responses ≤3.3°C	>3/8 rabbits show an increase of ≥0.5°C *AND/OR* the sum of all responses >3.3°C
21 CFR 610.13 (FDA 2005)	3	0/3 rabbits show an increase of ≥0.5°C	NA[1]
	5[1]	≤3/8 rabbits show an increase of ≥0.6°C *AND* the summed responses ≤3.7°C	>3/8 rabbits show an increase of ≥0.6°C *AND/OR* the summed responses >3.7°C
EP5.0 2.6.8 (EP 2005a)	3	Summed responses ≤1.15°C	Summed responses >2.65°C
	6[2]	Summed responses ≤2.80°C	Summed responses >4.30°C
	9[2]	Summed responses ≤4.45°C	Summed responses >5.95°C
	12	Summed responses ≤6.60°C	Summed responses >6.60°C
JP XIV (JP 2001)	3	3/3 rabbits show an increase of <0.6°C *AND* the summed responses ≤1.4°C	≥2/3 rabbits show an increase ≥0.6°C
	5[3]	≥4/5 rabbits show an increase <0.6°C	≥2/5 rabbits show an increase ≥0.6°C

CFR = U.S. Code of Federal Regulations; EP = European Pharmacopeia; FDA = U.S. Food and Drug Administration; JP = Japanese Pharmacopeia; NA = Not applicable; USP = United States Pharmacopeia; RPT = Rabbit pyrogen test

[1]If ≥1/3 rabbits show an increase of ≥0.5°C, continue test with an additional five rabbits.

[2]Three additional animals are tested when the summed responses falls in between the previous range.

[3]Five additional animals are tested when neither criterion is met, and results are based on these five animals only.

4.2 Reference Data Used to Assess *In Vitro* Test Method Accuracy

The ECVAM BRDs state that due to ethical and legal reasons, the RPT was not conducted in parallel to the *in vitro* test methods. Instead, historical RPT data produced over a 5-year period at the Paul-Ehrlich Institut (PEI), which is the German Federal Agency of Sera and Vaccines, were used (Hoffmann et al. 2005a). 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, neither the USP (USP 2007b) nor the EP (EP 2005a) prescribes a specific rabbit strain for the RPT.

4.3 Availability of Original Records for the *In Vivo* Reference Data

Section 4.1 of each ECVAM BRD indicates that the PEI provided the historical RPT data.

4.4 *In Vivo* Data Quality

The historical RPT studies were conducted at the 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 WHO collaborating center for QA of blood products and *in vitro* diagnostics. The unit for pyrogen and endotoxin testing of the PEI is accredited following ISO/IEC 17025 (International Standards Organization [ISO]

2005). In a request for additional information from ECVAM, it was stated that the RPT data was generated according to the EP monograph, but the detailed protocol used by this laboratory was not provided.

4.5 Availability and Use of Toxicity Information from the Species of Interest

A number of studies have concluded that humans and rabbits have approximately the same threshold to pyrogenic stimulation, although higher doses are more pyrogenic and more toxic in humans (Co Tui and Schrift 1942; Westphal 1956; Keene et al. 1961). Moreover, Greisman and Hornick (1969) compared three purified endotoxin preparations in rabbits and in male volunteers and showed that the threshold pyrogenic dose was similar in both species. However, the dose-response relationships for humans were considerably steeper than those for the rabbit at each dose tested.

As stated in **Section 1.2.1**, the major regulatory requirement for pyrogenicity testing is for end-product release of human and animal parenteral drugs, medical devices, and human biological products. The results from such testing are used to limit, to an acceptable level, the risks of febrile reactions from injection and/or implantation of the product of concern.

Endotoxin can produce a number of acute effects on human health. McKinney et al. (2006) reported increased cytokine expression patterns in a cohort of subjects experiencing systemic adverse events (i.e., fever, rash, lymphadenopathy) after smallpox vaccine administration. Martich et al. (1993) studied systemic, cardiovascular, pulmonary, cytokine release, and the inflammatory response resulting from i.v. injection of small doses of endotoxin in humans to understand mechanisms of sepsis and septic shock. Burrell (1994) later reviewed the available literature on the adverse human responses to bacterial endotoxin. In addition, environmental or chronic exposure to inhaled bacterial endotoxin (present in soil, in water, and on vegetation) may lead to an inflammation in the airways and/or gastrointestinal disturbances (Rylander 2002). Therefore, for protection of both human and animal health, it is vital that the test method employed provide an accurate estimation of the potential for a pyrogenic reaction.

4.6 Information on the Accuracy and Reliability of the *In Vivo* Test Method

Hoffmann et al. (2005a) modeled the sensitivity and specificity of the RPT using historical data (summarized in **Section 4.2**) to establish a threshold pyrogen dose (i.e., the endotoxin dose at which fever was induced in 50% of the rabbits). A threshold value of 0.5 EU/mL was defined by regression analysis of the data. The performance characteristics of the RPT (i.e., sensitivity and specificity) were then determined using a 2 x 2 contingency table, incorporating the parameters obtained from the regression analysis. The authors considered the prevalence of the endotoxin spikes included in the ECVAM accuracy evaluations in the validation studies (i.e., 0 EU/mL: 20%; 0.25 EU/mL: 20%; 0.5 EU/mL: 40%; 1.0 EU/mL: 20%) and applied the threshold pyrogen dose of 0.5 EU/mL to calculate theoretical values for sensitivity (58%) and specificity (88%) of the RPT.

The accuracy and reliability of the RPT for endotoxin testing has been considered adequate for U.S. and international regulatory needs for many years. Since its inclusion in the USP in 1941, the RPT has been used extensively and is the preferred method for detection of pyrogenicity for product development, because of the inability of the BET to detect non-endotoxin pyrogens.

[This Page Intentionally Left Blank]

5.0 Test Method Data and Results

5.1 Test Method Protocol

The Standard Operating Procedures (SOPs) used during the ECVAM validation studies are included in **Appendix A**. As described in **Section 2.1**, there are many similarities among the protocols for each of the *in vitro* pyrogen test methods, with very few notable differences other than the type of cells used (i.e., WB cells, PBMCs, monocytoid cell line) and the proinflammatory cytokine assayed (i.e., IL-1β or IL-6). These similarities and differences are outlined in **Table 2-1**. An internationally accepted endotoxin standard (i.e., WHO-LPS 94/580 [*E. coli* 0113:h10:K-]) was used to spike samples of saline or marketed parenteral pharmaceuticals. The same pharmaceuticals were used to create the spiked samples for all five test methods (see **Table 3-1** and **3-2**). These samples were included in a series of studies designed to determine the relevance and reliability of each of the *in vitro* pyrogen test methods.

5.2 Availability of Copies of Original Data Used to Evaluate Test Method Performance

ECVAM provided raw data from the validation studies in an electronic format (Excel® spreadsheets) that consisted of OD_{450} measurements for all replicates included in each of the validation studies.

NICEATM attempted to obtain additional *in vitro* and/or *in vivo* pyrogen test method data. A *Federal Register* (*FR*) notice (Vol. 70, No. 241, pp. 74833-74834, December 16, 2005) was published requesting original *in vitro* pyrogen test method and reference data from the currently used pyrogen test methods (i.e., RPT and/or BET). In addition, the *FR* notice was sent directly to more than 100 interested stakeholders internationally. Despite these efforts, no additional data were submitted.

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data

Details of the statistical approaches used to evaluate the accuracy and reliability of each of the five *in vitro* test methods are included in Section 5.3 of each ECVAM BRD. Briefly, as indicated in **Section 3.2**, 10 substances (each spiked with four concentrations of endotoxin, with one concentration spiked in duplicate) were tested in each test method to evaluate accuracy, while three substances (each spiked with three concentrations of endotoxin, with one tested in duplicate) were used to evaluate test method reproducibility. Varying concentrations of endotoxin-spiked saline were tested for the analysis of intralaboratory repeatability.

The evaluation of intralaboratory repeatability included coefficient of variation (CV) analysis of the log-transformed OD_{450} measurements for the replicates of each endotoxin concentration. Boxplots were also generated to demonstrate variability among these values for each concentration. Similar analyses were conducted for the three substances used to assess intra- and inter-laboratory reproducibility.

The reproducibility analysis incorporated the decision criteria that were developed to differentiate between pyrogenic and non-pyrogenic materials (using a threshold value of 0.5

EU/mL). In all reproducibility analyses, a single run consisted of each of the substances (as described above and in **Section 3.2**) assayed in quadruplicate. Acceptability criteria for each run included a CV analysis to remove highly variable samples from the analyses. This criterion ranged from a CV<0.25 to <0.45, depending on the test method being considered. For the measurement of intralaboratory reproducibility, pair-wise comparisons between the runs were determined and the associations between runs expressed as a percentage of agreement between two individual laboratories. It should be noted that this analysis takes into account the agreement of the resulting pyrogenicity decision (i.e., pyrogenic or non-pyrogenic), but does not consider whether the decision is correct. 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. Similar analyses were conducted for an assessment of interlaboratory reproducibility, in which pairwise comparisons between laboratories were determined and the associations were expressed as a percentage of agreement. This analysis included each run from each laboratory (n=3 per laboratory) and all possible interlaboratory combinations were compared. Similar to the intralaboratory analysis, this analysis takes the resulting pyrogenicity call from each run in each laboratory into consideration, but does not consider whether the call is correct. **Section 7.0** provides additional details and the resulting data from these analyses.

For the accuracy analysis, 2 x 2 contingency tables were constructed using the decision criteria defined in **Table 4-2** to assign a pyrogenicity call. Each run for each sample from each laboratory was considered independently. Accordingly, the *in vitro* call was compared to the "true status" (based on the known endotoxin spike concentration) of the sample. The resulting accuracy statistics were calculated based on the overall database for each test method. Similar to the reproducibility analyses, acceptability criteria for each run included a CV analysis to remove highly variable samples from the analyses, for which a range of CV<0.25 to CV<0.45 was used, depending on the test method being considered. **Section 6.0** provides additional details and the resulting data from these analyses.

Outliers were identified and eliminated using a two-step procedure. In the first step, replicates with an extremely large variation were identified by comparing the CV for the replicates with the extracted maximal CV (CV_{max}). If the CV for the replicates was smaller than the CV_{max}, then the data were analyzed without modification. However, if the replicates failed to pass this initial test, then the data were transformed with the natural logarithm and examined for outliers using the nonparametric Dixon's test (Dixon 1950; Barnett and Lewis 1984) or the Grubbs' test (Grubbs 1969) for normally distributed samples. If one observation was responsible for the large variation, then the observation was excluded. If the variation was due to all observations, then the entire set of replicates was excluded from further analysis. Additional information on the analytical procedure used to identify and eliminate outlier observations can be found in the materials provided by ECVAM (see **Appendix C**).

5.4 Summary of Results

Graphical representations of the repeatability and reproducibility analyses are provided in Section 5.2 of each ECVAM BRD (see **Appendix A**). The tabulated results from which the intra- and inter-laboratory reproducibility analyses and accuracy analyses can be conducted are provided in Section 5.4 of the ECVAM BRDs. The tables in that section include the test

substance name, the endotoxin spike concentration, the pyrogenicity call for each *in vitro* run, and the "true status" of each test substance.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained using coded chemicals and reported in accordance with GLP guidelines (i.e., OECD 1998; EPA 2003a, 2003b; FDA 2003). **Section 3.4** indicates that the 10 test substances and the four spike concentrations used for the accuracy evaluation were blinded to the testing laboratories. However, although the three spike concentrations were blinded to the participating laboratories for the reproducibility studies, the identity of the three test substances was not blinded.

5.6 Lot-to-Lot Consistency of Test Substances

Lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is used for the duration of the study. In these studies, the test substances were released from clinical lots of parenteral pharmaceuticals, which implied that they had been subjected to rigorous chemical manufacturing control analyses to verify that the compositions are consistent. However, the specific lot numbers for the test substances used in the validation study were not initially provided in the ECVAM BRDs. In response to a request for additional information, ECVAM provided this information (**Table 3-1** and **Appendix C**). In addition, the international standard for Gram-negative endotoxin, WHO-LPS 94/580 (*E. coli* O113:H10:K-), was used as the spike solution, which provides a measure of consistency for the positive control substance and the spike substance.

5.7 Availability of Data for External Audit

As described in **Section 8.4**, all records are stored and archived by the participating laboratories and are available for inspection.

[This Page Intentionally Left Blank]

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.

[This Page Intentionally Left Blank]

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.

8.0 Test Method Data Quality

8.1 Adherence to National and International GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines (i.e., OECD 1998; EPA 2003a, 2003b; FDA 2003). These guidelines provide an internationally standardized approach for the reporting requirements of studies designed for regulatory submissions, internal audits of laboratory records and data summaries, the archive of study data and records, and information about the test protocol and laboratory personnel, to provide assurances regarding the integrity, reliability, and accountability of the study.

The initial ECVAM validation studies for the five *in vitro* pyrogen test methods were conducted "in the spirit of" GLP requirements (i.e., written protocols and approved SOPs were followed during the entire course of the study). In the catch-up validation studies, two GLP laboratories and two National Control Laboratories participated.

8.2 Data Quality Audits

Formal assessments of data quality, such as a QA audit, generally involve a systematic and critical comparison of the data provided in a study report with the laboratory records generated for the study. No attempt was made to formally audit the quality of the data presented in the five ECVAM BRDs. However, as indicated in **Section 5.2**, the raw data from the validation studies are available from the participating laboratories for a quality analysis.

8.3 Impact of Deviations from GLP Guidelines

The impact of the deviations from the GLP guidelines, as reported in the ECVAM BRDs, was not evaluated.

8.4 Availability of Laboratory Notebooks or Other Records

All records are stored and archived by the participating laboratories and are available for inspection.

8.5 Need for Data Quality

Data quality is a critical component of the validation process. To ensure data quality, ICCVAM recommends that all data generated during the validation of a test method be available, along with the detailed protocol(s) under which the data were produced. Original data should be available for examination, as should supporting documentation such as laboratory notebooks. Ideally, the data should adhere to GLP guidelines (ICCVAM 1997). Data protocols for the validation studies summarized here are available from ECVAM (see **Appendix A**), and the data from the individual laboratories are available for inspection, as indicated in **Section 8.4**.

[This Page Intentionally Left Blank]

9.0 Other Scientific Reports and Reviews

The individual BRDs submitted by ECVAM (i.e., one for each of the *in vitro* pyrogen test methods) are provided in **Appendix A** and were used in the performance analyses described in **Section 6.0** and **Section 7.0**. A *FR* notice (Vol. 70, No. 241, pp. 74833-74834, December 16, 2005) was published requesting the submission of data from the RPT, the BET, or an *in vitro* pyrogen test method. No data were received in response to this request.

NICEATM conducted a prescreen evaluation of the ECVAM BRDs to verify that the information contained within the documents fulfilled the requirements outlined in the ICCVAM submission guidelines (ICCVAM 2003). Based on this evaluation, the ICCVAM PWG requested a direct comparison on the accuracy analysis of the *in vitro* test methods with the reference test methods (i.e., the RPT and the BET) and data to support the claim that the *in vitro* test methods can detect non-endotoxin pyrogens. In response to these requests, ECVAM provided supplemental data from published (e.g., Francois et al. 2006) and unpublished studies in an attempt to address these issues (see questions #1 and #2 in **Appendix B**).

Hartung et al. (2001) provided a summary report of an ECVAM-sponsored workshop to review the current status of pyrogenicity testing, to review the capabilities of new pyrogen tests, and to provide recommendations for their continued development. The need for alternatives to the RPT and the BET was discussed, and their respective limitations were highlighted. The workshop compared the utility of the various methods (i.e., *in vitro* pyrogen test methods, BET, RPT) for testing a variety of pyrogenic materials. Workshop conclusions indicated a need for alterative test methods to address the limitations of the BET and RPT, but stressed the need for appropriate validation of any new method.

9.1 Summaries of *In Vitro* Pyrogen Test Methods and Data from Published and Unpublished Studies

As indicated in **Section 1.5**, NICEATM conducted an online literature search for relevant information on the proposed test methods using multiple databases (i.e., PubMed, SCOPUS, TOXLINE, Web of Science). This search revealed ten additional scientific publications that contained data from *in vitro* pyrogen product testing. These studies contained comparisons of the results obtained in an *in vitro* test method with those obtained in the RPT and/or BET (see **Tables 9-1** to **9-8**). These studies were not included in previous sections of the ICCVAM BRD because they used a different method or protocol, or because they lacked sufficient information for an evaluation of accuracy and reliability (e.g., an adequate validation study design was not included, a standardized reference pyrogen was not used). Summaries of these published studies and available data from the *in vitro* pyrogen methods are presented below.

9.1.1 Andrade et al. (2003)

The authors evaluated the utility of human PBMCs and diluted WB for *in vitro* pyrogen tests and compared the responses to those obtained in the BET and RPT for the same diverse sampling of parenteral pharmaceuticals and biological products (see **Tables 9-1** and **9-2**). Interference testing of each substance was performed with spikes of the international endotoxin standard WHO-LPS 94/580. These studies established an endotoxin detection limit

of 0.06 EU/mL for both *in vitro* assays, and the results were consistent with those from the BET and RPT. The authors concluded that both the PBMC and WB methods were comparable to the BET and the RPT in their ability to detect and quantify the presence of endotoxin. In addition, the WB test method was able to detect concentration-dependent IL-6 release on exposure of WB to non-endotoxin pyrogens and pyrogens from Gram-positive organisms (i.e., *Candida albicans* and *Staphylococcus aureus*).

Table 9-1 Results of Pyrogen Testing of Pharmaceutical/Biological Products in the Human PBMC Assay, the BET, and the RPT[1]

Product	Number of Batches[2]	PBMC (EU/mL)	BET (EU/mL)	RPT
Ampicillin - 1000 mg/5 mL A	1	<6	<0.06	Pass
Ampicillin - 1000 mg/5 mL A	1	<6	<0.06	Pass
Gentamycin - 80 mg/2 mL	2	<3	<0.06	Pass
Oxacillin - 500 mg/5 mL	2	<3	<0.06	Pass
Enoxaparin - 100 mg/mL	3	<1.2	<0.06	Pass
Insulin - 100 U/mL	2	<3	<0.06	Pass
Tenoxican - 40mg/2 mL	1	<6	<0.06	Pass
Metoclopramide - 10 mg/2 mL	4	<3	<0.06	Pass
Calcium folinate - 50 mg/5 mL	1	<2.4	<0.06	Pass
Ranitidine - 25 mg/mL	2	<6	1.2-2.4	Pass
Pantoprazol - 40 mg/10 mL	1	<3	<0.06	Pass
Human serum albumin - 20%	1	<4.8	0.48-0.96	Pass
Erythropoietin - 4000 IU/vial A	1	<1.2	0.48-0.96	Pass
Erythropoietin - 2000 IU/vial B	1	112 ± 10[1]	491-983	Fail
Erythropoietin - 4000 IU/vial C	1	<1.2	<0.06	Pass
recG-CSF - 200 μg/vial A	3	<0.6	<0.06	Pass
Saline solution - 0.9% A	1	<0.3	<0.06	Pass

Abbreviations: BET = Bacterial Endotoxin Test; CSF = Colony Stimulating Factor; EU/mL = Endotoxin units/mL; IU = international units; PBMC = Peripheral blood mononuclear cells; rec = Recombinant; RPT = Rabbit pyrogen test; U = units

[1]From Andrade et al. (2003)

[2]Batch results were combined; PBMC and BET study values represent a mean ± standard deviation value or consensus detection limits (n=3 donors; 4 replicates from each donor).

Table 9-2 Results of Pyrogen Testing of Pharmaceutical/Biological Products by the Human WB Culture Assay, the BET, and the RPT[1]

Product	Number of Batches[2]	WB Culture (EU/mL)	BET (EU/mL)	RPT
Dipyrone - 500 mg/mL	3	<24	<0.06	Pass
Amikacin - 500 mg/2 mL	2	<12	<0.06	Pass
Ampicillin - 1000 mg/5 mL A	1	<6	<0.06	Pass
Ampicillin - 1000 mg/5 mL A	1	<6	<0.06	Pass
Gentamycin - 80 mg/2 mL	2	<6	<0.06	Pass
Oxacillin - 500 mg/5 mL	2	<6	<0.06	Pass
Vancomycin - 500 mg/5 mL	2	<6	<0.06	Pass
Enoxaparin - 100 mg/mL	3	<0.6	<0.06	Pass
Heparin - 5000 IU/mL	2	<0.6	<0.06	Pass
Insulin - 100 U/mL	3	<6	<0.06	Pass
Ketoprofen - 100 mg/2mL	1	<6	<0.06	Pass
Diclofenac - 75 mg/3 mL	1	<12	<0.06	Pass
Tenoxicam - 40 mg/2 mL	2	<6	<0.06	Pass
Metoclopramide - 10 mg/2 mL	3	<3	<0.06	Pass
Cytarabine - 100 mg/5mL	1	<1.2	<0.06	Pass
Calcium folinate - 50 mg/5 mL	1	<0.6	<0.06	Pass
Ranitidine - 25 mg/mL	1	<6	1.2-2.4	Pass
Pantoprazol - 40 mg/10 mL	1	<6	<0.06	Pass
Furosemide - 10 mg/mL	2	<0.6	<0.06	Pass
rec-hGH - 4 IU/vial A	2	<0.2	<0.06	Pass
rec-hGH - 4 IU/vial B	1	12.4±2.5[1]	15.84-31.68	Pass
Human serum albumin - 20%	1	<2.4	0.48-0.96	Pass
Erythropoietin - 4000 IU/vial A	1	0.76	0.48-0.96	Pass
Erythropoietin - 2000 IU/vial B	1	141±2.8[1]	491-983	Fail
Erythropoietin - 4000 IU/vial C	1	<0.6	<0.06	Pass
recG-CSF - 300 μg/vial	3	<0.6	<0.06	Pass
Saline solution 0.9% A	2	<0.3	<0.06	Pass
Saline solution 0.9% B	1	44.8±5[1]	48-96	Fail
Glucose - 0.5%	1	2054±95[1]	1920-3840	Fail
Vitamin K - 10 mg/mL	2	<6	<0.06	Pass

Abbreviations: BET = Bacterial endotoxin test; CSF = Colony stimulating factor; EU/mL = Endotoxin units/mL; hGH = Human growth hormone; IU = International units; rec = Recombinant; RPT = Rabbit pyrogen test; WB = Whole blood; U = units

[1]From Andrade et al. (2003)

[2]Batch results were combined; PBMC and BET study values represent a mean ± standard deviation value or consensus detection limits (n= 3 donors; 4 replicates from each donor).

9.1.2 Bleeker et al. (1994)

This study measured IL-6 release from PBMCs as an indicator of pyrogenicity for *in vitro* safety testing of hemoglobin (Hb) solutions. The authors demonstrated that pure, polymerized Hb produced under aseptic conditions did not induce or inhibit IL-6 production, whereas production under non-aseptic conditions led to IL-6 release, which was also seen with the BET. Based on these findings, the authors concluded that IL-6 release from isolated PBMCs provides a sensitive indicator of endotoxin contamination in Hb solutions. The observed detection limit for endotoxin in Hb solutions (below 0.4 EU/mL) led the authors to suggest that this test method would be more sensitive to the presence of endotoxin than the RPT.

9.1.3 Carlin and Viitanen (2003)

Using WB and MM6-based *in vitro* pyrogen methods, this study evaluated the pyrogenic potential of a multivalent vaccine, Infanrix® (GlaxoSmithKline) that contains protein and polysaccharide components from both Gram-positive and Gram-negative bacteria. The five Infanrix® vaccines studied (e.g., Infanrix®, Infanrix® Hep B, Infanrix® polio, Infanrix® hexa, and Infanrix® polio Hib) contain Gram-positive bacterial components that are potentially pyrogenic but not detectable in the BET. IL-6 production in the WB/IL-6 test method varied among the seven donor blood samples in response to each of the five vaccines. Some donor samples produced a weak or no IL-6 release and others produced a large release (**Table 9-3**). However, IL-6 production from any single donor was similar for all vaccines when tested at various times. The variability in the magnitude of response to each vaccine among donors and the consistency of the response of any single donor was also seen when IL-1β was used as a marker. IL-6 release from WB was also examined following exposure to three concentrations of endotoxin standard (0.2, 2, and 20 pg/1.2 mL). All donor WB samples released IL-6 in a concentration responsive manner.

The IL-6 release from MM6 cells (**Table 9-4**) exposed to the five Infanrix® vaccines was measured using an ELISA and compared to the responses induced by three concentrations of endotoxin standard (0.2, 2, and 20 pg/1.2 mL) in three separate experiments. The MM6 cells produced minimal responses to the vaccines when compared to WB, but released significant amounts of IL-6 in response to high concentrations of endotoxin. However, IL-6 induction by two different endotoxin standards in MM6 cells was strongly attenuated (>80% inhibition) when either of two vaccines (Infanrix® and Infanrix® Hep-B) was present (data not included in **Table 9-4**). Based on these studies, the authors suggested that a BET or RPT result might not correlate with the human fever response one might expect in humans immunized with such vaccines, because the production of proinflammatory cytokines may be compromised by various components in the vaccine product, and because Gram-positive components in the vaccines would not be detected in the BET.

Table 9-3 IL-6 Production from WB after Exposure to Endotoxin or Five Infanrix® Vaccines[1,2]

Experiment (Blood Donor)	Endotoxin		Vaccine (Absorbance in ELISA; n=4[3])					
	Endotoxin (pg/1.2 mL)	Absorbance (ELISA)	Dilution (µL vaccine/1.2 mL	Infanrix®	Infanrix® Hep-B	Infanrix® Hexa	Infanrix® Polio	Infanrix® Polio Hib
1	0.2	0.47	0.03	0.945	1.052	1.069	0.869	1.082
	2	0.971	0.3	1.826	2.055	2.014	1.832	1.919
	20	1.116	3	2.826	2.587	2.638	2.609	2.2
2	0.2	0.001	0.03	0.149	0.256	0.231	NT	0.284
	2	0.127	0.3	0.869	0.847	1.095	NT	0.933
	20	0.764	3	1.998	1.986	2.187	NT	1.685
3	0.2	-0.007	0.03	0.005	0.037	0.009	0.007	0.208
	2	0.09	0.3	0.275	0.457	0.282	0.321	0.261
	20	0.811	3	0.941	1.057	0.795	1.284	1.325
4	0.2	0.006	0.03	0.056	0.053	0.028	0.088	0.104
	2	0.043	0.3	0.165	0.312	0.44	0.309	0.533
	20	0.458	3	1.229	1.489	1.476	1.181	1.242
5	0.2	0.043	0.03	-	0.071	-0.003	-0.003	0.011
	2	0.024	0.3	0.007	0.014	0.004	0.03	0.05
	20	0.435	3	0.042	0.164	0.008	0.08	0.12
6	0.2	0.013	0.03	-0.009	-0.018	-0.01	-0.022	0.012
	2	0.022	0.3	-0.007	-0.008	0.005	-0.019	-0.007
	20	0.569	3	0.132	0.411	0.042	0.132	0.188
7	0.2	0.036	0.03	-0.012	-0.012	-0.01	-0.014	0.07
	2	0.014	0.3	-0.01	-0.01	-0.012	-0.011	-0.013
	20	0.436	3	0.183	0.274	0.045	0.183	0.525

Abbreviations: ELISA = Enzyme-linked immunosorbent assay; Hep = Hepatitis; IL-6 = Interleukin-6; NT = Not tested; WB = Whole blood

[1]From Carlin and Viitanen (2003)

[2]WB was challenged with endotoxin standard or vaccine in pyrogen-free water to provide the final concentration and incubated overnight at 37°C.

[3]Duplicate samples were run in two separate experiments.

Table 9-4 IL-6 Production by MM6 Cells after Exposure to Endotoxin or Five Infanrix® Vaccines[1,2]

MM6 Batch	Endotoxin			Vaccine (Absorbance in ELISA; 250,000 MM6 cells); n=4[3]				
	Endotoxin (pg/1.2 mL)	Absorbance in IL-6 ELISA	Dilution (µL vaccine/ 1.2 mL	Infanrix®	Infanrix® Hep-B	Infanrix® Hexa	Infanrix® Polio	Infanrix® Polio Hib
1	0.2	-0.001	0.3	0.013	0.014	0.001	0.002	-0.001
	2	0.026	3	0.078	0.158	0.06	0.105	0.07
	20	0.383	30	0.054	0.052	0.053	0.106	0.089
2	0.2	-0.001	0.3	0.004	0.01	0.001	0.003	0.004
	2	0.025	3	0.033	0.062	0.019	0.037	0.032
	20	0.4	30	0.013	0.012	0.018	0.038	0.038
3	0.2	-0.009	0.3	-0.012	-0.017	-0.021	-0.014	-0.019
	2	0.03	3	0.019	0.05	0.01	0.043	0.026
	20	0.192	30	-0.018	-0.012	-0.007	0	0.005

Abbreviations: ELISA = Enzyme-linked immunosorbent assay; IL-6 = Interleukin-6; MM6 = Mono Mac 6

[1]From Carlin and Viitanen (2003)

[2]MM6 cells were stimulated with endotoxin standard or vaccine in pyrogen-free water to provide the final concentration and incubated overnight at 37°C.

[3]n = Duplicate samples were run in two separate experiments.

9.1.4 Carlin and Viitanen (2005)

This study provides support for the findings from a previous study (Carlin and Viitanen (2003) in which the authors demonstrated IL-6 release by a WB method in response to pyrogenic or spiked multivalent vaccine preparations that were inactive in the BET. It also confirms that IL-6 was released from WB of some, but not all donors. The present study demonstrates that IL-6 release in susceptible donors was caused by toxoids from Gram-positive diphtheria, and to a lesser extent, from tetanus bacterial components of the vaccines. The WB donors were studied for two years and their responses to the individual vaccines, whether responsive or non-responsive, were consistent. The responses of these donors to Gram-negative endotoxin or lipoteichoic acid (LTA) from Gram-positive bacteria were consistent and confirmed the findings of Fennrich et al. (1999) with respect to the consistency of responses among several hundred blood donors to endotoxin. The authors concluded that individual donor-specific differences in IL-6 release from WB exposed to the multivalent vaccines resulted from toxoids present in the diphtheria or tetanus component, and noted that these donor-specific responses to the vaccines were not observed in the BET.

9.1.5 Daneshian et al. (2006)

This study describes the development of a modification to the WB/IL-1β method termed AWIPT (Adsorb, Wash, *In Vitro* Pyrogen Test). The authors indicate that this modification is intended to increase sensitivity to the presence of endotoxin contamination by isolating endotoxin from WB. To accomplish this, the sample containing endotoxin (naturally occurring or spiked) is treated with human serum albumin (HSA) covalently linked to macroporous acrylic beads. The HSA-treated beads bind the endotoxin, which is subsequently eluted from the beads. The WB/IL-1β test method is therefore performed using a slightly modified protocol in which the diluted WB is incubated overnight with the sample in the bead suspension.

The results showed that HSA-coated beads bind endotoxin in a concentration-dependent manner (when spiked with 0, 25, 50, and 100 pg/mL LPS), but little or none was bound to unmodified beads. The test showed a detection limit of 25 pg/mL LPS (i.e., 0.25 EU/mL), which is less sensitive than the BET (3 pg/mL) and more sensitive than the RPT (50 pg/mL). IL-1β secretion in response to either LPS or LTA was generally higher using the AWIPT procedure, but the concentrations of LPS or LTA needed to induce a response were similar; thus the sensitivity of this test modification was comparable to that of the unmodified WB/IL-1β test method.

Daneshian et al. studied the kinetics of cytokine release from WB in response to a challenge with 2 pg/mL of endotoxin. IL-1β release in the AWIPT-treated samples lagged slightly behind that of the standard WB/IL-1β test in the 0 to 8 hr time period, whereas more IL-1β was produced in the AWIPT-treated samples in the 10 to 30 hr time period. Some immunomodulatory or toxic cancer drug samples tested in the WB/IL-1β method interfered with the WB/IL-1β assay and required a higher dilution (1/10 to 1/100) to detect IL-1β. Detection of endotoxin spiked into these test samples (measured as IL-1β release) generally occurred at lower dilutions in AWIPT than in the WB/IL-1β test method, suggesting that the interfering substances were removed by the procedure. For example, five dilutions (ranging from 1/3 to 1/316) of liposomal daunorubicin were spiked with 25 pg/mL of endotoxin and

detection of IL-1β was compared between the two methods. This cytokine was not detectable in the WB/IL-1β method (< 30% of the IL-1β released by endotoxin) at any drug dilution, whereas in the AWIPT, IL-1β was detected at drug dilutions of 1/32, 1/100, and 1/316 (>78% of the IL-1β released by endotoxin).

The authors concluded that the inclusion of endotoxin adsorption and washing steps in the WB/IL-1β method (i.e., the AWIPT) to remove potentially interfering substances improved the detection of pyrogenic contaminants in immunomodulatory and toxic cancer drug samples. They suggest that the AWIPT method offers an improvement for safety testing of products administered to patients, and for batch control in pharmaceutical processing.

9.1.6 Eperon et al. (1996, 1997)

Eperon and colleagues developed an *in vitro* test system for measuring pyrogenic substances using two clones derived from MM6 cells (Professor Ziegler-Heitbrock, University of Munich) and one from a THP-1 cell line (European Collection of Animal Cell Cultures, Porton Down, Salisbury, U.K.). These clones are reported to be more phenotypically stable over time with respect to their superior responsiveness to endotoxin than the parent cell lines. Endotoxin content was measured by the release of TNF-α using an immunoassay. These clones demonstrate high LPS sensitivity when non-pyrogenic fetal calf serum is used in the assay as a serum supplement. Enhanced expression of the cell-surface endotoxin receptor CD14 was obtained by pretreatment of the cells for two days with calcitrol. Purified endotoxin (i.e., LPS; smooth strain and rough mutant), other cellular components from Gram-negative or Gram-positive bacteria, and Mycobacteria were tested. The MM6 clones responded to these pyrogenic products in an order of potency of detection equivalent to that found in the RPT and similar to that observed in the BET (i.e., Gram-negative endotoxin > Gram-positive material > non-endotoxin pyrogens). The response of the THP-1 clone was similar to that of the MM6 clones, except that the THP-1 clone did not respond to diphosphoryl lipid A, a structural component of LPS.

Pyrogen testing of a panel of stable blood products, including albumin and Immunoglobulin G (IgG) for parenteral use, produced similar results in the RPT and MM6 or THP-1 clones when tested as received (i.e., free of detectable pyrogens). The products produced positive results when spiked with 20 EU/mL of endotoxin (**Table 9-5**), with a few exceptions. For example, in the cell-based test, there was one borderline but significantly positive result in an unspiked sample, representing a false positive result relative to the RPT. In the BET, 4 of 13 (31%) unspiked samples tested positive (i.e., false positive). The results suggest that the cell-based assays may produce fewer false positives than the BET.

When 10 bacterial and viral vaccine preparations were evaluated, the monocytoid cell-based test method (e.g., combined results from two experiments with each cell line) correlated well with the RPT (positive or negative for endotoxin) with the exception of one preparation that produced nearly 10-fold less TNF-α than the other samples, and was near the limit of detection. This result was not significantly different from the negative control (**Table 9-6**). The authors suggest that these cloned monocytoid cell-based test methods are valid *in vitro* alternatives for detection of endotoxin in commercial preparations, and produce results comparable to the RPT and BET.

Table 9-5 Pyrogenic Activity of Blood Preparations for Parenteral Use[1]

Preparation	Endotoxin Spike	RPT[2]	BET[3,4]	Cell Test[5,6]
IgG for i.v. use	20 EU/mL	+	+	+
	20 EU/mL	+	+	+
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
Albumin	20 EU/mL	+	+	+
	20 EU/mL	+	+	+
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
	-	-	-	-
Test threshold[7]		*Δt = 1.5 °C*	*300 pg/mL LPS*	*50 pg/mL TNF*

Abbreviations: BET = Bacterial endotoxin test; EU/mL = Endotoxin units/mL; IgG = Immunoglobulin G; i.v. = Intravenous; LPS = Lipopolysaccharide; RPT = Rabbit pyrogen test; TNF = Tumor Necrosis Factor

[1]From Eperon et al. (1997)

[2]n=3

[3]n=2

[4]Haemachem BET (St. Louis)

[5]n=4 [Note: Cell type not specified; author claims that the Mono Mac 6 or acute monocyte leukemia THP-1 cell lines are equally capable of endotoxin detection.]

[6]TNF induction was determined using a commercial TNF Enyme-linked immunosorbent assay.

[7]RPT threshold was obtained from the European Pharmacopeia; the threshold for the BET and cell-based test methods was considered to be equal to 2 standard deviations from the mean of a set of negative samples.

Table 9-6 Pyrogenic Activity of Vaccine Preparations[1]

Vaccine Preparation[2]	Batch[3]	Cell Test[4]	Pyrogenicity
IgG for i.m. use	A-1	n.d.[5]	-
	A-2	n.d.	-
Bacterial vaccines	B-1	10.8±0.3	+
	C-1	6.0±3.6	+
	D-1	1.4±1.8	-
Viral vaccines	E-1	n.d.	-
	E-2	n.d.	-
	F-1	n.d.	-
	F-2	n.d.	-
	G-1	21.2±3.2	+

Abbreviations: IgG = Immunoglobulin G; i.m. = Intramuscular; n.d. = Non-detectable
[1]From Eperon et al. (1997)
[2]Vaccine solutions were tested at 1/20 (v/v)
[3]Letters refer to distinct types of vaccine preparations; numbers to different lots
[4]Tumor necrosis factor (TNF)-α production in ng/mL±standard error of the mean (n=3) [Note: Cell type not specified; author claims that the Mono Mac 6 or acute monocyte leukemia THP-1 cell lines are equally capable of endotoxin detection.]
[5]No measurable quantity of cytokine was detected.

9.1.7 Marth and Kleinhappl (2002)

In 2000, Ticovac®, a thiomersal- and albumin-free tick-borne encephalitis (TBE) vaccine, was developed as a more immunogenic alternative to previous vaccines that also produced fewer side effects. Although the Austrian health authorities approved this vaccine, 779 cases of fever were reported in children less than 15 years of age, including a high incidence of febrile convulsions in children ages 2 and younger. To determine the cause of these fever reactions, Ticovac® was compared to FSME-Immun®, a TBE-vaccine that rarely resulted in febrile reactions, in an *in vitro* human WB assay that measured cytokine release (i.e., IL-1β, IL-6, IL-8, and TNF-α) as an indication of immune system activation. Ticovac®, which differs from FSME-Immun® only in the albumin component, induced high amounts of TNF-α ($P \leq 0.0001$) and lower amounts of IL-1β ($P \leq 0.05$) as compared to FSME-Immun®. The addition of 0.5 mg of albumin (i.e., the identical quantity of albumin in FSME-Immun®) to Ticovac® reduced the TNF-α induction significantly, resulting in TNF-α production that was similar to the level stimulated by FSME-Immun®. The incubation of Ticovac® with human WB resulted in an increase in TNF-α concentration after 4 hr (peaking at 15 hr) and returned to baseline levels by 27 hr. IL-1β release displayed a similar time course. This temporal response to Ticovac® correlated well with the progression of the clinical outcome (i.e., fever and convulsions in children 6 to 8 hr after the first immunization). Although the mechanism of cytokine production by Ticovac® is unknown, it is clearly linked to the absence of albumin, which is needed as a stabilizer to bind to the antigen of the vaccine. Thus, it was recommended that albumin be added to subsequently produced TBE-vaccines to inhibit nonspecific, excessive immunological reactions.

9.1.8 Martis et al. (2005)

The goal of this study was to establish the cause of 186 cases of aseptic peritonitis that occurred between 2001 and 2003 in peritoneal dialysis patients using an icodextrin-containing dialysate that met both European and USP standards. These patients were not febrile or toxic in appearance, but abdominal pain that was modest to absent and cloudy dialysate were common features. The authors conducted physical, chemical, and microbiological analyses on the recalled dialysate and calculated dose-response curves for IL-6 production in PBMCs from human donors and for sterile peritonitis in rats. Increased levels of IL-6 were identified in dialysis solutions of compliant batches (n=3), but not in non-complaint batches (n=2). Effluents from compliant batches also stimulated IL-6 release in the PBMC assay. Polymyxin B did not inhibit this response, suggesting that a lipopolysaccharide was not responsible for the increased IL-6 levels. When neither Gram-negative nor Gram-positive bacterial contamination was identified in the dialysates, the possibility of a non-endotoxin contaminant was considered. A Gram-positive bacterial cell wall component (i.e., peptidoglycan [PG]) was identified using a silkworm larvae assay in a significant number of dialysates. In the PBMC assay, IL-6 release increased with PG concentration in a dose-response manner. A microbial investigation revealed that the dialysates were contaminated with a Gram-positive organism (i.e., *Alicyclobacillus acidocaldarius*), which contains approximately 40% PG in its cell wall. In rat studies, intraperitoneal injection of icodextrin containing PG (0 - 5000 μg/mL) produced a dose-dependent inflammatory response as measured by an increase in TNF-α and IL-6 production. Subsequent PG contamination has been eliminated using more stringent filtration and carbon treatment steps in the manufacturing process, assaying for PG contamination with the silkworm larvae test, and measuring IL-6 production with the PBMC assay. The lack of aseptic peritonitis incidents that have occurred since the implementation of these additional detection processes support the concept that PG contamination of dialysate was responsible for the reported cases of aseptic peritonitis.

9.1.9 Pool et al. (1998)

This study describes a WB assay for the detection of pyrogens in blood products. IL-6 release from WB in response to endotoxin is used to define a pyrogenic response. This assay was highly responsive to *E. coli* endotoxin (i.e., the limit of detection of endotoxin was 1.25 EU/mL), and also responded to whole bacteria (*E. coli* and *Bacillus subtilis*). There was considerable variation in IL-6 levels released from WB between donors following exposure to endotoxin, but each donor response was always linear. The potential pyrogenicity of production batches of HSA, fibronectin (Fn), and stabilized human serum (SHS) solutions were evaluated using the WB method and compared to the BET and RPT. Spike recovery in batches of these samples varied between 90 and 116% for *E. coli* endotoxin, 74 to 111% for *B. subtilis*, and 61 to 99% for *E. coli* and the products tested did not interfere with the IL-6 assay system. Good correlations were found among the WB, BET, and RPT results (**Table 9-7**). Of 22 products tested, the WB assay and the RPT were in agreement (i.e., pass or fail) for all tests, while one sample was classified as negative in the BET, but positive in both the WB method and the RPT. The detection limit for endotoxin by the WB method was 1.25 EU/mL, which is lower than the established pyrogen cut-off level (as stated in the European Pharmacopeia) for the products under investigation (i.e., 2 EU/mL for HSA and SHS; 4.5

EU/mL for Fn). The authors concluded that the WB assay was able to detect both Gram-negative and a Gram-positive pyrogens and exhibited greater sensitivity to endotoxin than the RPT.

Table 9-7 Comparison of the WB test, BET, and the RPT for Detecting Pyrogens in Production Batches of Biological Products[1]

Product	Batch	WB (EU/mL)[2]	BET	RPT
Fibronectin - 0.5 mg/mL	Fn3195	<0.05	Pass	Pass
	Fn3296	<0.05	Pass	Pass
	Fn3596	1.28	Pass	Pass
Human serum albumin - 200 mg/mL	B274	29.4	Fail	Fail
	B291	<0.05	Pass	Pass
	B293	<0.05	Pass	Pass
	B294	<0.05	Pass	Pass
	B295	<0.05	Pass	Pass
	B296	<0.05	Pass	Pass
	B297	<0.05	Pass	Pass
	B298	1	Pass	Pass
	B299	1.1	Pass	Pass
	B300S	1	Pass	Pass
	B301	<0.05	Pass	Pass
	B302	>20	Pass[3]	Fail
Stabilized human serum - 50 mg/mL	SS349	0.7	Pass	Pass
	SS350	<0.05	Pass	Pass
	SS351	<0.05	Pass	Pass
	SS352	0.5	Pass	Pass
	SS353	<0.05	Pass	Pass
	SS354	0.6	Pass	Pass
	SS355	0.5	Pass	Pass

Abbreviations: BET = Bacterial endotoxin test; EU/mL = Endotoxin units/mL; RPT = Rabbit pyrogen test; WB = Whole blood

[1]From Pool et al. (1998)

[2]Result based on interleukin-6 secretion in human WB using an enzyme-linked immunosorbent assay calibrated to an *E. coli* endotoxin standard (Kabi Diagnostica).

[3]False negative relative to the RPT response

9.1.10 Taktak et al. (1991)

This paper summarizes the development of an *in vitro* pyrogen test method based on IL-6 release from MM6 cells. A detectable level of IL-6 was released in response to 2.5 pg/mL of endotoxin, yielding a level of sensitivity of 25 pg/mL when testing 5% HSA at a 1/10 dilution for the presence of endotoxin. Three batches of a therapeutic HSA that caused fever in humans were positive in the MM6/IL-6 method, whereas the same substances were negative in the BET and the RPT (**Table 9-8**). As in the BET, the samples required a 1/10 dilution to remove interfering substances. The assay had sensitivity equal to that of the BET (25 pg/mL) and 40-fold greater than the RPT (1000 pg/mL). The authors suggest that the MM6/IL-6 method represents an important alternative to the existing pyrogen tests and may be a more appropriate end-product test for the detection of pyrogens in parenteral products, such as HSA, that cannot be detected in the BET.

Table 9-8 Results of Pyrogen Testing of Batches of Therapeutic HSA Using the MM6/IL-6, BET, and RPT[1]

Batch of HSA	Endotoxin Quantitation by IL-6 Release (pg/mL)	Endotoxin Quantitation by the BET		RPT Result
		IU/mL	pg/mL[6]	
1[2]	97±2.3[3,4]	1.0-2.0	140-280	Pass
2[2]	30±2.8[4]	2.4-3.2	336-448	Pass
3[2]	31±2.3[4]	0.5-0.75	70-105	Pass
4	<25[5]	<0.24	<34	Pass
5	<25[5]	3.6-4.8[7]	504-762	Pass
6	<25[5]	<0.26	<36	Pass

Abbreviations: BET = Bacterial endotoxin test; HSA = Human serum albumin; IL-6 = Interleukin-6; IU = International units; MM6 = Mono Mac 6; RPT = Rabbit pyrogen test

[1]From Taktak et al. (1991)

[2]Batch of HSA used that caused fever in humans.

[3]Mean ± standard error of the mean

[4]Values are significantly different from subthreshold concentrations of endotoxin (<2.5 pg/mL; $p<0.001$).

[5]Values below the detection limit of the test system (25.0 endotoxin units/mL); preparations of HSA were tested at a dilution of 1/10, and 2.5 pg/mL endotoxin was the lowest concentration of endotoxin tested that evoked a significant release of IL-6.

[6]1.0 IU=0.14 ng for preparation used.

[7]False positive relative to RPT.

9.2 Conclusions from Scientific Literature Based on Independent Peer-Reviewed Reports and/or Reviews

An additional nine reports describing studies of cell-based *in vitro* pyrogen methods were obtained from the literature search described in **Section 1.5**. Although these reports did not include data on test substances that could be used in the performance analysis in **Section 6.0** and **Section 7.0**, they did evaluate the use of the *in vitro* pyrogen test methods for sensitivity to endotoxin (i.e., endotoxin detection limit), specificity of the response to endotoxin and/or non-endotoxin pyrogens (i.e., spectrum and relative potency of various pyrogens detected), and/or the impact of interfering substances. However, they did not compare results from the *in vitro* test methods to results from the RPT, BET, or human fever reaction. A summary of each study is presented below.

9.2.1 De Groote et al. (1992)

The authors measured the release of various cytokines (IL-1β, IL-6, TNF-α, IL-2, IFN- γ , and granulocyte-macrophage colony stimulating factor [GM-CSF]) in response to endotoxin or phytohemagglutinin (PHA) stimulation of WB and PBMC cultures. Endotoxin stimulated IL-1β, TNF-α, and IL-6 release, while PHA stimulated IL-2, IFN- γ , and GM-CSF release. There was a significant correlation between production of the three endotoxin-induced cytokines and the number of monocytes in the challenged culture, suggesting that monocytes are the major source of these cytokines: the other cytokines did not correlate with any of the cell types. The data also suggested that WB produced less variable levels of cytokines than PBMC on exposure to endotoxin. Consistent results were obtained with the WB test using more than 50 different blood donors. The authors suggest that WB is a more appropriate choice for studying cytokine production *in vitro* and its modulation by exogenous or endogenous factors, because natural cell-to-cell interactions are preserved, immune mediators are available, and cytokine levels obtained with PBMC were more variable.

9.2.2 Fennrich et al. (1999)

Fennrich and colleagues compared a commercially available human WB/IL-1β pyrogen assay (PyroCheck® from DPC Biermann, Bad Nauheim distributed by Millenia, U.K.) to the BET and RPT. There was a concentration-dependent IL-1β release in WB that was incubated with nitrocellulose filters containing live *E. coli* bacteria and *E. coli* killed by heat or by antibiotics. The authors also tested air conditioning filters from a veterinary sheep facility and identified filters to be contaminated with bacteria that were later confirmed by microbiological tests (the data and the identity of the organisms identified were not presented). The authors compared the PyroCheck®, BET, and the RPT (**Table 9-9**) and concluded that PyroCheck® is a simple, accurate test that detects a wider range of pyrogens than the BET.

Table 9-9 Comparison of the Application Spectra of the RPT, the BET, and the Human WB Assay (PyroCheck®)[1]

Test		Applications		
		PyroCheck®	BET	RPT
Pyrogens	Gram-negative	+	+	+
	Gram-positive	+	-	+
	Fungi	+	-	+
Product pyrogenicity	Biologicals	+	-	+
	Pharmaceuticals	+	+	+
	Medical devices	+	+[2]	-
	Air quality	+[2]	+[2]	-
	Blood products	+	-	-

Abbreviations: BET = Bacterial endotoxin test; RPT = Rabbit pyrogen test; WB = Whole blood
[1]From Fennrich et al. (1999)
[2]Based on preliminary data

9.2.3 Hansen and Christensen (1990)

This study compared the results from PBMC exposed to endotoxin or ultraviolet light-killed *S. aureus* as an index of pyrogenicity, and then compared these results to the BET and the RPT. The authors used human PBMC obtained from heparinized peripheral blood and measured IL-1-like material in culture supernatants by evaluating co-mitogenic activity on PHA-stimulated murine thymocytes (measured in units of IL-1β where 1 unit is defined as the concentration that gives 50% of the maximal incorporation of ^{3}H-thymidine in the thymocyte assay). The endpoint is referred to as an IL-1-like material because other cytokines such as IL-2, IL-6, and TNF-α may also stimulate the proliferative response of the thymocytes. When exposed to endotoxin, PBMC secreted cytokines in a concentration-dependent manner that provided a limit of detection of 200 pg/mL of endotoxin. In comparison, the BET can normally detect 10 to 100 pg/mL of endotoxin, while the RPT can detect 500 pg/mL. Therefore, the PBMC procedure had a level of detection of endotoxin 2.5-fold lower than that of the RPT and 2-fold higher than the BET. The PBMCs also responded with greater sensitivity to the Gram-positive pyrogen *S. aureus* (10^5 cells/mL), which was not detected in the BET (10^9 cells/mL). Based on these results, the authors proposed that the PBMC test be used as an alternative *in vitro* test to the BET and RPT.

9.2.4 Hartung and Wendel (1996)

The authors stimulated human WB with various inflammatory agents to release endogenous cytokines (i.e., IL-1β, TNF-α) and inflammatory mediators (i.e., prostaglandin E_2) as an *in vitro* method for the detection of pyrogenic materials. Cytokines were released in a concentration-dependent manner following exposure to endotoxin or LTA. Heat-killed Gram-positive bacteria (*S. aureus*) or components of these organisms (i.e., muropeptides, LTA, enterotoxins, streptolysin O) and plant mitogens such as phorbol myristate acetate and PHA also produced a cytokine response. Higher concentrations (three orders of magnitude) of the Gram-positive pyrogens were needed to elicit a response as compared to Gram-negative pyrogenic material.

Studies to determine the variability among the responses of different donor WB samples were also performed. Only two of the 18 donor samples released IL-1β in response to 1 pg/mL of endotoxin, but all responded to 10 pg/mL endotoxin. The release of IL-1β from the WB samples of 45 individual donors exposed to 100 ng/mL of endotoxin was also consistent. Based on these results, the authors suggested using the WB/IL-1β test method as an *in vitro* alternative to the RPT.

9.2.5 Moesby et al. (1999)

Moesby and colleagues compared pyrogen testing using MM6 cells, isolated PBMC, and the BET. LPS and ultraviolet (UV) light-killed Gram-negative *Staphylococcus typhimurium* or Gram-positive *S. aureus* produced concentration-dependent increases in IL-6 production in MM6 or PBMC cultures. PBMC, but not MM6 cells, were able to differentiate UV-irradiated yeast (*C. albicans*) and mold (*Aspergillus niger*) pyrogens, as evidenced by statistically significant increases in IL-6 production. The BET can detect Gram-negative endotoxin, but not Gram-positive endotoxin or LTA (the pyrogenic component of Gram-positive bacteria),

and it may weakly detect yeast or viral pyrogens that the MM6 assay could not detect. Therefore, the authors suggest that pyrogen testing using MM6 cells would be a useful supplement to the BET for the detection of both Gram-negative and Gram-positive bacteria.

9.2.6 Nakagawa et al. (2002)

Nakagawa and colleagues describe an *in vitro* pyrogen test system based on proinflammatory cytokine release from a sub-clone of MM6 cells (i.e., MM6-CA8) and compare this response to a human WB culture system and the RPT. Similar to MM6 cells, MM6-CA8 were developed for superior reactivity to both endotoxin and PG. The MM6-CA8 cells release IL-6, TNF-α, and IL-1, but in greater quantities than MM6 cells in the range of 1 to 1000 pg/mL of endotoxin (up to 4-fold greater) or to 1 to 1000 ng/mL PG (up to 10-fold greater) compared to MM6 cells. The range of responses of human WB to the various pyrogens was similar to that of the MM6-CA8 cells. The relative potencies of the various pyrogens in the RPT were similar to those of the cytokine-induction potencies in the WB and MM6-CA8 methods, except for polyinosinic:polycytidylic acid, which was reported to be 10,000-fold more potent as a pyrogen injected in rabbits when compared to humans. The authors conclude that these results suggest MM6-CA8 cells can detect a variety of pyrogens using IL-6 as the marker, and that these responses are highly relevant to the prediction of human fever reactions.

9.2.7 Pool et al. (1999)

This article describes a method to differentiate between endotoxin and non-endotoxin pyrogens when testing HSA solutions in a WB culture assay. Detection limits for four Gram-positive (*Bacillus stearothermophilus*, *B. subtilis*, *Micrococcus luteus*, and *S. aureus*) and four Gram-negative bacteria (*E. coli*, *Kleibsiella pneumoniae*, *Proteus vulgaris*, and *Pseudomonas aeruginosa*) were expressed as the number of whole bacteria required to produce a pyrogenic response equal to that of 1.25 EU/mL endotoxin. *B. stearothermophilus* and *E. coli* produced concentration-dependent increases in IL-6 production. The cationic antibiotic Polymyxin B, which inhibits the binding of endotoxin to the CD14 receptor, produced concentration-dependent inhibition of IL-6 release following exposure to 10 EU/mL endotoxin in the WB assay at concentrations up to 1 EU/mL and completely inhibited IL-6 release at concentrations above 2 EU/mL. In contrast, Polymyxin B had no effect on IL-6 release following exposure to *B. subtilis*. These data suggest that Polymyxin B may be useful for differentiating endotoxin and non-endotoxin pyrogenic contaminants. The data also suggests that binding of endotoxin to Polymyxin B (e.g., by linkage to an affinity column) may be used in the depyrogenation process.

9.2.8 Poole et al. (2003)

This paper describes a rapid single-plate *in vitro* test for the presence of pyrogenic substances based on monocyte activation. The assay uses polyclonal antibodies to IL-6 or TNF-α cytokines, coated and stabilized onto 96-well plates. Monocytoid cells (e.g., PBMC, MM6 or THP-1 cells), endotoxin standard (LPS), test sample, and a second biotinylated antibody specific for the cytokine (e.g., either IL-6 or TNF-α) are incubated for 2 to 4 hr in the antibody-coated wells. An ELISA for one of the cytokines is then performed on the washed plate. IL-6 is preferred and provides a limit of detection of 0.015 EU/mL with PBMC, 0.05

EU/mL in MM6 cells, and 0.03 EU/mL with diluted WB. The amount of TNF-α released in WB in response to endotoxin was approximately 50 to 70% lower than IL-6, but was released earlier (i.e., 2 vs. 4 hr). The amount of IL-6 released on exposure to endotoxin tended to be greater in this single plate test when compared to the traditional two-plate test (i.e., in which the supernatant from one plate is transferred to a second plate for the ELISA) using PBMCs, MM6 cells, THP-1 cells, or WB. The authors report that this single plate assay using IL-6 release as the endpoint can be completed in 5 hr, and that this time could be reduced to 3 hr using TNF-α as the endpoint (because it is released earlier from the cells). The authors also suggest that this single plate test method is readily adaptable to high-throughput assays.

9.2.9 Schindler et al. (2004)

The authors optimized conditions for use of cryopreserved human WB in pyrogen testing to obviate the need for fresh WB. The release of IL-1β from fresh and Cryo WB collected from five donors was used as the measure of endotoxin presence. Challenge with 0.5 or 1.0 EU/mL endotoxin resulted in IL-1β release in bloods from all donors, although kinetic studies suggested that IL-1β release was delayed one hr in the cryopreserved samples. Cryopreservation did not appear to alter the spectrum of detectable pyrogens or immune stimuli when results were compared to that of fresh WB, and no cytokine release was measured in materials that fresh WB did not respond to. Seven clinical-grade (i.e., endotoxin-free) parenteral products spiked with 0.5 EU/mL of endotoxin revealed that there was less interference in Cryo WB than in fresh WB based on lower minimal interference dilutions that were always at or below the MVD for each product. The data showed that a broad variety of drugs could be tested for pyrogenic contaminants using Cryo WB while maintaining the ELC established in the various Pharmacopeias.

[This Page Intentionally Left Blank]

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement)

10.1 How the Five *In Vitro* Test Methods Will Refine, Reduce, or Replace Animal Use

ICCVAM promotes the scientific validation and regulatory acceptance of new methods that refine, reduce, or replace animal use where scientifically feasible. Refinement, Reduction, and Replacement are known as the three "Rs" of animal alternatives. These principles of humane treatment of laboratory animals are described as:

- Refining experimental procedures such that animal suffering is minimized
- Reducing animal use through improved science and experimental design
- Replacing animal models with non-animal procedures (e.g., *in vitro* technologies), where possible (Russell and Burch 1959)

In 2002, a total of 243,838 rabbits were used in the U.S. for all research and testing purposes, of which 6,324 rabbits were reported as experiencing more than slight or momentary pain and/or distress where anesthetics, analgesics, or tranquilizers could not be administered for scientific reasons (USDA 2002). Eight of these cases were specifically attributed to pyrogenicity testing, presumably based on induction of a fever response (USDA 2002). Thus, although the potential for more than slight or momentary pain and/or distress exists for pyrogenicity testing when a fever response is induced, it does not appear that a fever response is common. In 2006, a total of 239,720 rabbits were used in the U.S. for all research and testing purposes (USDA 2006). No data related to pyrogenicity testing were reported.

In Canada, a total of 18,152 rabbits were used for all scientific purposes in 2006, 3,485 of which were used for regulatory studies and the development of products (Canadian Council on Animal Care [CCAC] 2007). Although no specific data for the number of animals used for pyrogenicity testing were reported, it is likely that the number of rabbits used for this purpose is less than the total of 3,485 used for both regulatory studies and product development.

In the EU, approximately 313,000 total rabbits were used for all scientific purposes in 2005 (Commission of the European Communites [CEC] 2007). Of these, approximately 276,000 rabbits were used for pharmaceutical products and medical device testing (i.e., either research and development, production and quality control, or toxicological and other safety evaluations). Although the number of rabbits specifically used for pyrogenicity testing was not reported, it is likely that this number is significantly less than the total of 276,000.

In the U.K., a total of 21,736 procedures (which used 14,712 total rabbits due to reuse of some test animals) were performed using rabbits for all scientific purposes in 2004 (Home Office 2005). Of these procedures, 8,488 were specifically attributed to pyrogenicity testing in rabbits. Although the total number of rabbits used for these procedures were not provided, it is likely less than 8,488 rabbits based on the assumption that some animals were reused. In 2006, a total of 20,378 procedures (which included 13,397 total rabbits) were performed in the U.K. for all scientific purposes (Home Office 2007). No specific data for pyrogenicity testing were reported in 2006.

The currently accepted pyrogen test methods require the use of rabbits or horseshoe crab hemolymph. The proposed *in vitro* pyrogen test methods address each aspect of animal welfare outlined above. These assays 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.

10.2 Requirement for the Use of Animals

10.2.1 Rationale for the Use of Animals

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, humans are the animals used for these assays. 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 appropriately treated.

11.0 Practical Considerations

Several issues are taken into account when assessing the practicality of using an *in vitro* test method in place of an *in vivo* test method. In addition to reliability and accuracy evaluations, assessments of the laboratory equipment and supplies needed to conduct the *in vitro* test method, level of personnel training, labor costs, and the time required to complete the test method relative to the *in vivo* test method are necessary. The time, personnel cost, and effort required to conduct the proposed test method(s) must be considered to be reasonable when compared to the *in vivo* test method it is intended to replace.

11.1 Transferability of the *In Vitro* Pyrogen Test Methods

Test method transferability addresses the ability of a method to be accurately and reliably performed by multiple laboratories (ICCVAM 2003), including those experienced in the particular type of procedure as well as laboratories with less or no experience in the particular procedure. The degree of transferability of a test method can be evaluated by its interlaboratory reproducibility. ECVAM measured the transferability (i.e., interlaboratory reproducibility) of each assay among experienced laboratories. The results presented in **Tables 7-3** and **7-4** provide an estimate of the minimum variability to be expected. Interlaboratory variability is anticipated to be greater (i.e., lower transferability) among laboratories that have less experience with the assays.

11.1.1 Facilities and Major Fixed Equipment

A standard laboratory facility for sterile tissue culture is necessary for performing the *in vitro* pyrogen test methods. The major equipment necessary to conduct the tests are readily available and include a laminar flow hood, tissue culture incubator, water bath, and spectrophotometric microplate reader.

In contrast, the RPT requires a facility that meets applicable State and Federal regulations for the care and housing of laboratory animals. The primary expense for equipping a facility to conduct the RPT would be the acquisition of an adequate animal room and associated housing (e.g., cages, bedding, food, water, etc.) for boarding animals during the study.

11.1.2 General Availability of Other Necessary Equipment and Supplies

The equipment and supplies necessary to conduct the *in vitro* pyrogen test methods (e.g., micropipetters, sterile tissue culture vessels, disposable plastic ware, assay reagents) are readily available in most scientific laboratories, or can be obtained from any of several scientific laboratory equipment vendors.

The RPT requires fewer general laboratory supplies. Those that are needed are readily available in most laboratories, or could be readily obtained from any of a number of scientific laboratory equipment vendors.

11.2 Personnel Training Considerations

Training considerations are defined as the level of instruction needed for personnel to conduct the test method accurately and reliably (ICCVAM 2003). Evaluation of the levels of

training and expertise needed to conduct the test method, as well as the training requirements needed to insure that personnel are competent in the test procedures, are discussed below.

11.2.1 Required Training and Expertise Needed to Conduct the In Vitro Pyrogen Test Methods

Laboratory personnel require training with the relevant ELISA procedures and the aseptic techniques associated with mammalian tissue culture. The quality criteria associated with each *in vitro* test method may be used to ensure that personnel are competent in the performance of the various procedures. When a technician has mastered all aspects of the protocol, and can independently conduct the assay such that the quality criteria have been met, the individual is considered to have demonstrated proficiency in the assay.

The RPT requires training in the care and handling of laboratory animals, and the collection of accurate rectal temperature measurements at the appropriate time intervals from each rabbit. The laboratory personnel must be adequately trained to maintain the animals, and to accurately and consistently record the proper body temperature. It is not known what, if any, proficiency requirements are in place for the RPT.

11.3 Cost Considerations

In addition to the major fixed equipment and overhead requirements, three additional factors contribute to the overall cost of the proposed *in vitro* test methods: 1) cost and licensing fees associated with the MM6 monocytoid cell line, 2) cost of the reagents for the ELISA procedure, and 3) personnel costs associated with obtaining human blood and performing the test methods. With respect to the RPT, the direct and indirect costs of operating an animal facility must be considered. The most notable expenses will likely include personnel to care for the maintenance of the rabbits, staff to perform the RPT, and veterinarians to monitor the health of the rabbits. As summarized in **Table 11-1**, cost estimates from various contract laboratories that perform the RPT or from one contract laboratory that performs an ELISA-based *in vitro* pyrogen test using human WB indicate that the *in vitro* test methods are considerably more cost effective (i.e., by about a factor of ten) than the RPT. Furthermore, the use of high throughput procedures to analyze the *in vitro* pyrogen tests may provide further reduced costs per test substance.

11.4 Time Considerations

The *in vitro* pyrogen methods require two half-days (i.e., one before and one after the overnight incubation) to complete if cryopreserved blood or MM6 cells are available. If fresh WB is used or if interference testing is needed, additional time will be required. On the first day, the test materials are prepared and incubated with the monocytoid cells. On the second day, cytokine release from the cells is determined by an ELISA procedure. The BET and RPT can both be completed within one working day. However, according to the USP30 NF25<151> (USP 2007b) procedure for the RPT, each rabbit must be conditioned prior to its first use by a sham test that includes all steps of pyrogenicity testing except for injection.

Table 11-1 Cost Estimates for the RPT and *In Vitro* Pyrogen Tests

Contract Laboratory	Test or Cell Line	GLP Compliant	Cost Estimate per Test	Additional Information
A	RPT	Yes	$2100[1]	-
B	RPT	Yes	$4050[1]	-
C	RPT	Yes	$3600[1]	-
D	IPT/HumanWB	ND	$315[2]	Cost decreases with number of test substances; $315 per 1 test substance; $210 per 2 to 10 test substances; $105 per 11 or more test substances. Note: IPT is not a licensed product and should not be used for the release of drugs.
E	MM6	NA	Negotiable	Use of MM6 cells for product testing require negotiation of a fee for provision and a royalty payment per batch of product tested with Dr. HWL Ziegler-Heitbrock at the University of Leicester, Dept of Microbiology, Leicester, U.K.

Abbreviations: GLP = Good laboratory practice; IPT = *In vitro* pyrogen test; MM6 = Mono Mac 6; NA = Not applicable; ND = Not determined; RPT = Rabbit pyrogen test; WB = Whole blood

[1]Each RPT includes one test substance, one positive, and one negative control performed in triplicate. Thus, a minimum of 9 rabbits is needed per test.

[2]Each IPT includes one test substance, one positive, and one negative control performed in triplicate.

[This Page Intentionally Left Blank]

12.0 References

Andrade SS, Silveira RL, Schmidt CA, Junior LB, Dalmora SL. 2003. Comparative evaluation of the human whole blood and human peripheral blood monocyte tests for pyrogens. Int J Pharm 265:115-124.

Barnett V, Lewis T. 1984. Outliers in statistical data. In: Wiley Series in Probability and Mathematical Statistics. (Barnett V, Lewis T, eds). 3rd ed. New York:John Wiley and Sons.

Beutler B, Rietschel ET. 2003. Innate immune sensing and its roots: the story of endotoxin. Nat Rev Immunol 3:169-176.

Bleeker WK, de Groot EM, den Boer PJ, Biessels PT, Asrden LA, Bakker JC. 1994. Measurement of interleukin-6 production by monocytes for *in vitro* safety testing of hemoglobin solutions. Artif Cells Blood Substit Immobil Biotechnol 22:835-840.

Brunson KW, Watson DW. 1974. Pyrogenic specificity of streptococcal exotoxins, staphylococcal enterotoxin, and gram-negative endotoxin. Infect Immun 10:347-351.

Burrell R. 1994. Human responses to bacterial endotoxin. Circ Shock 43:137-153.

Carlin G, Viitanen E. 2003. *In vitro* pyrogenicity of a multivalent vaccine; infanrix. Pharmeuropa 15:418-423.

Carlin G, Viitanen E. 2005. *In vitro* pyrogenicity of the diphtheria, tetanus and acellular pertussis components of a trivalent vaccine. Vaccine 23:3709-3715.

CCAC. 2007. CCAC Survey of Animal Use 2006. Ottawa, Ontario:Canadian Council on Animal Care (CCAC). Available: http://www.ccac.ca.

CEC. 2007. Report from the Commission to the Council and the European Parliament. Fifth Report on the Statistics on the Number of Animals Used for Experimental and Other Purposes in the Member States of the European Union (Annex). Brussels:Commission of the European Communities (CEC). Available: http://ec.europa.eu/environment/chemicals/lab_animals/pdf/5th_stat_rep_lab_animals_en.pdf

Conti B, Tabarean I, Andrei C, Bartfai T. 2004. Cytokines and fever. Front Biosci 9:1433-1449.

Cooper JF, Levin J, Wagner HN. 1971. Quantitative comparison of *in vitro* and *in vivo* methods for the detection of endotoxin. J Lab Clin Med 78:138-148.

Co Tui, Shrift, MH. 1942. Alternative test for pyrogens in infusion fluids. Proc Soc Exp Biol Med 49:320-330.

Daneshian M, Guenther A, Wendel A, Hartung T, von Aulock S. 2006. *In vitro* pyrogen test for toxin or immunomodulatory drugs. J Immunol Methods 313:169-175.

De Groote D, Zangerle PF, Gevaert Y, Fassotte MF, Beguin Y, Noizat-Pirenne F, et al. 1992. Direct stimulation of cytokines (IL-1β, TNF-α, IL-6, IL-2, IFN- γ and GM-CSF) in whole blood. I. Comparison with isolated PBMC stimulation. Cytokine 4:239-248.

Dinarello CA, O'Conner JV, LoPreste G, Swift RL. 1984. Human leukocytic pyrogen test for detection of pyrogenic material in growth hormone produced by recombinant Escherichia Coli. J Clin Microbiol 20:323-329.

Dinarello CA. 1999. Cytokines as endogenous pyrogens. J Infect Dis 179:S294-304.

Dinarello CA. 2004. Infection, fever, and exogenous and endogenous pyrogens: some concepts have changed. J Endotoxin Res 10:201-222.

Ding JL, Chai C, Pui AWM, Ho B. 1997. Expression and full length and deletion homologues of *Carcinoscorpus rotundicauda* Factor C in *Saccharomyces cerevisciae*: immunoreactivity and endotoxin binding. J Endotoxin Res 4:33-43.

Ding JL, Ho B. 1998. Cloned Factor C cDNA of the Singapore horseshoe crab *Carsinoscorpius rotundicauda* and purification of Factor C proenzyme. U.S. Patent No. 5 716 834.

Ding JL, Ho B. 2001. A new era in pyrogen testing. Trends Biotechnol 19:277-281.

Dixon WJ. 1950. Analysis of extreme values. Ann Math Stat 21:488-506.

Duff GW, Atkins E. 1982. The detection of endotoxin by *in vitro* production of endogenous pyrogen: comparison with *Limulus* amebocyte lysate gelation. J Immunol Methods 52:323-331.

EP. 2005a. Biological Tests. Supplement 5.2. 2.6.8 Pyrogens. In: European Pharmacopoeia (EP). 5th Ed. Strasbourg, France:Council of Europe - European Directorate for the Quality of Medicines.

EP. 2005b. Biological Tests. Supplement 5.2. 2.6.14 Bacterial Endotoxins. In: European Phamacopoeia (EP). 5th Ed. Strasbourg, France:Council of Europe - European Directorate for the Quality of Medicines.

EPA. 2003a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 792.

EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40 CFR 160.

Eperon S, Jungi TW. 1996. The use of human monocytoid lines as indicators of endotoxin. J Immunol Methods 194(2):121-129.

Eperon S, de Groote DD, Werner-Felmayer G, Jungi TW. 1997. Human monocytoid cell lines as indicators of endotoxin: comparison with rabbit pyrogen and *Limulus* amoebocyte lysate assay. J Immunol Methods 207:135-145.

FDA. 2003. Good Laboratory Practices for Nonclinical Laboratory Studies. 21 CFR 58.

FDA. 2005. General Biological Products Standards: General Provisions: Purity. 21 CFR 610.13.

Fennrich S, Wendel A, Hartung T. 1999. New applications of the human whole blood pyrogen assay (PyroCheck). ALTEX 16:146-149.

Francois C, Neveu J, Sauvaire D, Bonnet PA, Tissier MH. 2006. *In vitro* pyrogenicity of Gram-positive bacteria - validation of the kit using fresh human whole blood. Pharmeur Sci Notes 2006:17-21.

Gaines Das RE, Brügger P, Patel M, Mistry Y, Poole S. 2004. Monocyte activation test for pro-inflammatory and pyrogenic contaminants of parenteral drugs: test design and data analysis. J. Immunol Methods 288:165-177.

Greisman SE, Hornick RB. 1969. Comparative pyrogenic reactivity of rabbit and man to bacterial endotoxin. Proc Soc Exp Biol Med 131:1154-1158.

Grubbs FE. 1969. Procedures for detecting outlying observations in samples. Technometrics 11:1-21.

Hansen EW, Christensen JD. 1990. Comparison of cultured human mononuclear cells, *Limulus* amebocyte lysate and rabbits in the detection of pyrogens. J Clin Pharm Ther 15:425-433.

Hartung T, Wendel A. 1996. Detection of pyrogens using human whole blood. In Vitro Toxicol 9:353-359.

Hartung T, Aaberge I, Berthold S, Carlin G, Charton E, Coecke S, et al. 2001. Novel pyrogen tests based on the human fever reaction. The report and recommendations of ECVAM Workshop 43. Altern Lab Anim 29:99-123.

Hochstein HD, Fitzgerald EA, McMahon FG, Vargas R. 1994. Properties of US standard endotoxin (EC-5) in human male volunteers. J Endotoxin Res 1:52-56.

Hoffmann S, Peterbauer A, Schindler S, Fennrich S, Poole S, Mistry Y, et al. 2005a. International validation of novel pyrogen tests based on human monocytoid cells. J Immunol Methods 298:161-173.

Hoffmann S, Luderitz-Puchel U, Montag T, Hartung T. 2005b. Optimisation of pyrogen testing in parenterals according to different pharmacopoeias by probabilistic modelling. J Endo Res 11:25-31.

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjostrom M. 1996. Recommendations of the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. Altern Lab Anim 24:511-530.

Home Office. 2005. Statistics of Scientific Procedures on Living Animals 2004. London:The Stationery Office. Available: http://www.official-documents.gov.uk/document/cm67/6713/6713.pdf

Home Office. 2007. Statistics of Scientific Procedures on Living Animals 2006. London:The Stationery Office. Available: http://www.homeoffice.gov.uk/rds/pdfs07/spanimals06.pdf

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No: 97-3981. Research Triangle Park, NC:National Toxicology Program.

ICCVAM. 2000. ICCVAM Authorization Act of 2000. Public Law 106-545 (42 U.S.C. 285).

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC:National Toxicology Program.

Ishii KJ, Coban C, Akira S. 2005. Manifold mechanisms of toll-like receptor-ligand recognition. 2005. J Clin Immunol 25:511-521.

Ishii KJ, Akira S. Innate immune recognition of, and regulation by, DNA. 2006. Trends Immunol. 27:525-532.

ISO. 2005. General Requirements for the Competence of Testing and Calibration Laboratories (Second edition). ISO/IEC 17025. Geneva:International Organization for Standardization (ISO).

ISO. 2006. Biological Evaluation of Medical Devices - Part 11: Tests for Systemic Toxicity (First edition 1993-12-15). ISO 10993-11. Geneva:International Organization for Standardization (ISO).

JP. 2001. Bacterial Endotoxins Test. In: The Japanese Pharmacopoeia 14th ed. (JP XIV) Part I. Shibuya, Shiubya-ku, Tokyo, Japan:The Society of Japanese Pharmacopoeia.

Keene WR, Silberman HR, Landy M. 1961. Observations on the pyrogenic response to and its application to the bioassay of endotoxin. J Clin Invest 40:295-301.

Marth E. and Kleinhappl B. 2002. Albumin is a necessary stabilizer of TBE-vaccine to avoid fever in children after vaccination. Vaccine 20:532-537.

Martich GD, Boujoukos AJ, Suffredini AF. 1993. Response of man to endotoxin. Immunobiology 187:403-416.

Martin MA, Roberts S. 2006. An evaluation of bootstrap methods for outlier detection in least squares regression. J Appl Stat 33:703-720.

Martis L, Patel M, Giertych J, Mongoven J, Tammine M, Perrier MA, et al. 2005. Aseptic perontonitis due to peptidoglycan contamination of pharmacopoeia standard dialysis solution. Lancet 365:588-594.

McKinney BA, Reif DM, Rock MT, Edwards KM, Kingsmore SF, Moore JH, Crowe Jr JE. 2006. Cytokine expression patterns associated with systemic adverse events following smallpox immunization. J Infect Dis 194:444-453.

Moesby L, Jensen S, Hansen EW, Christensen JD. 1999. A comparative study of Mono Mac 6 cells, isolated mononuclear cells and *Limulus* amoebocyte lysate assay in pyrogen testing. Int J Pharm 191:141-149.

Moesby L, Hansen EW, Christensen JD. 2000. Endotoxin testing of proteins for parenteral administration using the Mono Mac 6 assay. J Clin Pharm Ther 25:283.

Moesby L, Hansen EW, Christensen JD, Høyer CH, Juhl GL, Olsen HB. 2005. Dry and moist heat sterilization cannot inactivate pyrogenicity of Gram positive microorganisms. Eur J Pharm Sci 26:318-323.

Nakagawa Y, Maeda H, Murai T. 2002. Evaluation of the *in vitro* pyrogen test system based on proinflammatory cytokine release from human monocytes: comparison with a human

whole blood culture test system and with the rabbit pyrogen test. Clin Diagn Lab Immunol 9:588-597.

Netea MG, Kallberg BJ, Van der Meer JWM. 2000. Circulating cytokines as mediators of fever. Clin Inf Dis 31:S178-S184.

Norata GD, Garlaschelli K, Ongari M, Raselli S, Gigore L, Benenuto F, Maggi FM, Catapano AL. 2005. Effect of the Toll-like receptor 4 (TLR-4) variants on intima-media thickness and monocyte-derived macrophage response to LPS. J Intern Med 258:21-27.

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD Principles on Good Laboratory Practice (as revised in 1997). ENV/MC/CHEM(98)17. Paris:Organisation for Economic Co-operation and Development (OECD).

Park JM, Ng VH, Maeda S, Rest RF, Karin M. 2004. Anthrolysin O and other Gram-positive cytolysins are Toll-like receptor 4 agonists. J Exp Med 200:1647-1655.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. 1998. The detection of pyrogens in blood products using an ex vivo whole blood culture assay. J Immunoassay 19:95-111.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. 1999. Differentiation between endotoxin and non-endotoxin pyrogens in human albumin solutions using an *ex vivo* whole blood culture assay. J Immunoassay 20:79-89.

Poole S, Mistry Y, Ball C, Gaines Das RE, Opie LP, Tucker G, et al. 2003. A rapid 'one-plate' *in vitro* test for pyrogens. J Immunol Methods 274:209-220.

Poole S, Thorpe R, Meager A, Gearing AJ. 1988. Assay of pyrogenic contamination in pharmaceuticals by cytokine release from monocytes. Dev Biol Stand 69:121-123.

Russell WMS, Burch RL. 1959. The Principles of Humane Experimental Technique. Potters Bar, Herts, England:Universities Federation for Animal Welfare.

Rylander R. 2002. Endotoxin in the environment—exposure and effects. J Endotoxin Res 8:241-252.

Saper CB, Breder CD. 1994. The neurological basis of fever. New England J Med 330:1880-1886.

Schindler S, Asmus S, von Aulock S, Wendel A, Hartung T, Fennrich S. 2004. Cyropreservation of human whole blood for pyrogencity testing. J Immunol Methods 294:89-100.

Schindler S, Spreitzer I, Löschner B, Hoffmann S, Hennes K, Halder M, et al. 2006. International validation of pyrogen tests based on cryopreserved human primary blood cells. J Immunol Methods 316:42-51.

Stoll LL, Denning GM, Weintraub NL. 2006. Endotoxin, TLR4 signalling and vascular inflammation: potential therapeutic targets in cardiovascular disease. 12:4229-4245.

Taktak YS, Selkirk S, Bristow AF, Carpenter A, Ball C, Rafferty B, et al. 1991. Assay of pyrogens by interleukin-6 release from monocytic cell lines. J Pharm Pharmacol 43:578-582.

Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T, Tada K. 1980. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). Int J Cancer 26:171-176.

USDA. 2002. Animals Used in Research 2002. Available: http://www.aphis.usda.gov/animal_welfare/downloads/awreports/awreport2002.pdf

USDA. 2006. Animals Used in Research 2006. Available: http://www.aphis.usda.gov/animal_welfare/downloads/awreports/awreport2006.pdf

USP. 2007a. Bacterial Endotoxins Test. USP30 NF25<85>. Rockville, MD:The U.S. Pharmacopeial (USP) Convention.

USP. 2007b. Pyrogen Test. USP30 NF25<151>. Rockville, MD:The U.S. Pharmacopeial (USP) Convention.

van Deventer SJH. 2000. Cytokine and cytokine receptor polymorphisms in infectious disease. Intensive Care Med 26:S98-S102.

von Aulock S, Schroder NWJ, Gueninzius K, Traub S, Hofffmann S, Graf K, Dimmeler S, Hartung T, Schumann RR, Hermann C. 2003. Heterozygous Toll-like receptor 4 polymorphism does not influence lipopolysaccharide-induced cytokine release in human whole blood. J Infect Dis 188:938-943.

Walls EA, Berkson J, Smith SA. 2002. The horseshoe crab, *Limulus polyphemus*: 200 million years of existence, 100 years of study. Rev Fish Sci 10:39-73.

Westphal O. 1956. In: Polysaccharides in Biology (GF Springer, ed.). Trans. 2nd Conf. New York:Josiah Macy Jr. Foundation. 115.

Ziegler-Heitbrock HW, Thiel E, Futterer A, Herzog V, Wirtz A, Riethmuller G. 1988. Establishment of a human cell line (Mono Mac 6) with characteristics of mature monocytes. Int J Cancer 41:456-461.

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. Accuracy is a meaure of test method performance and one aspect of relevance. The term is often used interchangeably with concordance (see two-by-two table). Accuracy is highly dependent on the prevalence of positives in the poulation being examined.

Amebocytes: The blood cells of the horseshoe crab (*Limulus polyphemus* or *Tachypleus tridentatus*) that contain the active components of the reagent used in the BET.

Assay[2]: The experimental system used. Often used interchangeably with "test" and "test method."

Bacterial endotoxin test (BET)[3]: A test used to quantify endotoxins of Gram-negative bacterial origin using amebocyte lysate from the horseshoe crab. Two types of techniques exist: the gel-clot techniques, which are based on gel formation and the photometric techniques. The photometric techniques include the turbidimetric technique, which is based on the development of turbidity after cleavage of an endogenous substrate and a chromogenic method, which is based on the development of color after cleavage of a synthetic peptide-chromogen complex.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Coefficient of variation (CV): A statistical representation of the precision of a test. It is expressed as a percentage and is calculated as follows:

$$\left(\frac{standard\ deviation}{mean}\right) \times 100\%$$

Concordance[2]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with accuracy (see two-by-two table). Concordance is highly dependent on the prevalence of positives in the population being examined.

Endogenous pyrogens: Various cytokines including interleukins (e.g., IL-1α, IL-1β), tumor necrosis factor (i.e., TNF-α, TNF-β), and interferon (IFN-γ) released from leukocytes in response to external stimuli (e.g., endotoxin) capable of causing an increase in body temperature above the normal level.

---

[1]The definitions in this Glossary are restricted to the RPT, the *in vitro* pyrogen test methods included in this BRD, and the BET.
[2]From ICCVAM (2003)
[3]From USP (2005)

Endotoxin limit concentration (ELC): The concentration at which endotoxin is considered to be pyrogenic. It is expressed as the ratio of the threshold pyrogen dose (K) and the RPT dose or the maximum human dose administered on a weight (kg) basis in 1 hr (M) defined as K/M. The ELC varies based on M.

- The U.S. Food and Drug Administration (FDA) ELC for non-intrathecal medical devices is 0.5 EU/mL.
- The FDA ELC for intrathecal medical devices is 0.06 EU/mL.

Endpoint[2]: The biological or chemical process, response, or effect assessed by a test method.

False negative[2]: A substance incorrectly identified as negative by a test method.

False negative rate[2]: The proportion of all positive substances falsely identified by a test method as negative (see two-by-two table). It is one indicator of test method accuracy.

False positive[2]: A substance incorrectly identified as positive by a test method.

False positive rate[2]: The proportion of all negative substances that are falsely identified by a test method as positive (see two-by-two table). It is one indicator of test method accuracy.

Fever: Elevation of body temperature above the normal level.

Good laboratory practices (GLP)[2]: Regulations promulgated by the FDA and the U.S. Environmental Protection Agency, principles and procedures adopted by the Organization for Economic Cooperation and Development, and Japanese authorities that describe record keeping and QA procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Hazard[2]: The potential for an adverse health or ecological effect. A hazard potential occurs only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Interlaboratory reproducibility[2]: A measure of whether different qualified laboratories using the same protocol and test substances can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes and indicates the extent to which a test method can be transferred successfully among laboratories.

Intralaboratory repeatability[2]: The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period.

Intralaboratory reproducibility[2]: The first stage of validation; a determination of whether qualified people within the same laboratory can successfully replicate results using a specific test protocol at different times.

In vitro: In glass. Refers to assays that are carried out in an artificial system (e.g., in a test tube or petri-dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or purified cellular components.

In vivo: In the living organism. Refers to assays performed in multi-cellular organisms.

Lipopolysaccharide (LPS): A complex of lipid and carbohydrate (i.e., endotoxin) released from the cell walls of Gram-negative organisms that is pyrogenic and capable of producing septic shock.

Lipoteichoic acid: A polyol phosphate polymer bearing a strong negative charge that is covalently linked to the peptidoglycan in Gram-positive bacteria. It is strongly antigenic, but is generally absent in Gram-negative bacteria. Therefore, it is considered the primary pyrogenic component of Gram-positive bacteria.

Minimum valid concentration (MVC): The concentration of a product when it is diluted to the MVD expressed as λM/K, where:

- **λ =** The sensitivity of the *Limulus* Amebocyte Lysate (LAL) reagent used expressed as EU/mL. The value varies with the method employed. For the gel-clot method, it is the labeled LAL sensitivity (EU/mL). For the chromogenic, turbidometric, or kinetic-turbidometric methods, it is the lowest point used in the standard curve.
- **M =** The maximum human dose for pyrogenicity administered on a weight basis (kg) in 1 hr, or the RPT dose (whichever is larger). It is one of the variables used to define the ELC defined as the ratio of K/M, where K is the threshold pyrogen dose in rabbits or humans.
- **K** = See threshold pyrogen dose.

Maximum valid dilution (MVD): When a U.S. Pharmacopeia (USP) ELC is defined, the MVD is the ratio of the product of the ELC and the product potency to the LAL reagent sensitivity (λ) expressed as ([ELC x Product Potency]/λ). If there is no official USP ELC defined, then the MVD is the ratio of the Product Potency/MVC.

Monocytoid cells: Cells obtained from peripheral blood or grown in culture that phenotypically resemble monocytes or macrophages.

Negative control: An untreated sample containing all components of a test system, except the test substance solvent, which is replaced with a known non-reactive material, such as water. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Negative predictivity[2]: The proportion of correct negative responses among substances testing negative by a test method (see two-by-two table). It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Parenteral: Introduction into the body by some means other than through the gastrointestinal tract; referring particularly to intravenous (i.v.), intramuscular, subcutaneous, or intrathecal injection.

Performance[2]: The accuracy and reliability characteristics of a test method (see accuracy and reliability).

pH: A measure of the acidity or alkalinity of a solution. A pH of 7.0 is neutral; higher pHs are alkaline, lower pHs are acidic.

Positive control: A sample containing all components of a test system and treated with a substance known to induce a positive response, which is processed with the test substance-treated and other control samples to demonstrate the sensitivity of each experiment and to allow for an assessment of variability in the conduct of the assay over time.

Positive predictivity[2]**:** The proportion of correct positive responses among substances testing positive by a test method (see two-by-two table). It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

Prevalence[2]**:** The proportion of positives in the population of substances tested (see two-by-two table).

Protocol[2]**:** The precise, step-by-step description of a test method, including a list of all necessary reagents and criteria and procedures for evaluation of the test data.

Pyrogen: A substance that causes a rise in body temperature above normal or that produces a fever. Gram-negative, Gram-positive, and acid-fast bacteria, molds, viruses, and yeast and some of their cellular constituents are pyrogenic.

Quality assurance (QA)[2]**:** A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures is assessed independently by individuals other than those performing the testing.

Rabbit pyrogen test (RPT)[3]**:** A test designed to limit to an acceptable level the risks of febrile reaction in the patient to the administration, by injection, or the product concerned. The test involves measuring the rise in temperature of rabbits following the i.v. injection of a test solution.

Reduction alternative[2]**:** A new or modified test method that reduces the number of animals required.

Reference test method[2]**:** The accepted *in vivo* test method used for regulatory purposes to evaluate the potential of a test substance to be hazardous to the species of interest.

Refinement alternative[2]**:** A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Relevance[2]**:** The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the accuracy or concordance of a test method.

Reliability[2]**:** A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[2]**:** A new or modified test method that replaces animals with non-animal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[2]**:** The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

Sensitivity[2]**:** The proportion of all positive substances that are classified correctly as positive in a test method. It is a measure of test method accuracy (see two-by-two table).

Specificity[2]**:** The proportion of all negative substances that are classified correctly as negative in a test method. It is a measure of test method accuracy (see two-by-two table).

Test[2]**:** The experimental system used; often used interchangeably with "test method" and "assay."

Test method[2]**:** A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay" (see validated test method and reference test).

Test method component: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures.

Threshold pyrogen dose: The dose level at which a product is considered to be pyrogenic or non-pyrogenic. It is one of the variables (K) used to calculate the ELC defined as K/M, where M is the RPT dose or the maximum human dose administered in 1 hr (whichever is larger).

- The threshold pyrogen dose for non-intrathecal use in rabbits and humans is 5.0 EU/kg.
- The threshold pyrogen dose for intrathecal use in rabbits and humans is 0.2 EU/kg.

Tiered testing: A testing strategy where all existing information on a test substance is reviewed, in a specified order, prior to *in vivo* testing. If the irritancy potential of a test substance can be assigned, based on the existing information, no additional testing is required. If the irritancy potential of a test substance cannot be assigned, based on the existing information, a step-wise animal testing procedure is performed until an unequivocal classification can be made.

Transferability[2]**:** The ability of a test method or procedure to be accurately and reliably performed in different, competent laboratories.

Two-by-two table[2]**:** The two-by-two table can be used for calculating accuracy (concordance) ([a+d]/[a+b+c+d]), negative predictivity (d/[c+d]), positive predictivity (a/[a+b]), prevalence ([a+c]/[a+b+c+d]), sensitivity (a/[a+c]), specificity (d/[b+d]), false positive rate (b/[b+d]), and false negative rate (c/[a+c]).

		NEW TEST OUTCOME		
		Positive	**Negative**	**Total**
Reference Test Outcome	**Positive**	a	c	a + c
	Negative	b	d	b + d
	Total	a + b	c + d	a + b + c + d

Validated test method[2]**:** An accepted test method for which validation studies have been completed to determine the relevance and reliability of this method for a specific proposed use.

Validation[2]**:** The process by which the reliability and relevance of a procedure are established for a specific purpose.

Weight of evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.

Appendix A

ECVAM BRDs and Standard Operating Procedures

The following documents are available on request from NICEATM:

A1 **The Human Whole Blood (WB)/Interleukin (IL)-1β *In Vitro* Pyrogen Test**

A2 **The Human WB/IL-1β *In Vitro* Pyrogen Test: Application of Cryopreserved (Cryo) Human WB**

A3 **The Human WB/IL-6 *In Vitro* Pyrogen Test**

A4 **The Human Peripheral Blood Mononuclear Cell (PBMC)/IL-6 *In Vitro* Pyrogen Test**

A5 **The Monocytoid Cell Line Mono Mac 6 (MM6)/IL-6 *In Vitro* Pyrogen Test**

[This Page Intentionally Left Blank]

Appendix B

ECVAM Response to ICCVAM Questions

This document is available on request from NICEATM.

[This Page Intentionally Left Blank]

Appendix C

Additional Information Requested by the Panel

[This Page Intentionally Left Blank]

Appendix C1

ESAC Statement on the Validity of *In Vitro* Pyrogen Tests

[This Page Intentionally Left Blank]

EUROPEAN COMMISSION
DIRECTORATE GENERAL JRC
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

STATEMENT ON THE VALIDITY OF IN-VITRO PYROGEN TESTS

At its 24th meeting, held on 20-21 March 2006 at the European Centre for the validation of alternative methods (ECVAM), Ispra, Italy, the non-Commission members of the ECVAM Scientific Advisory Committee (ESAC)[1] unanimously endorsed the following statement:

Following a review of scientific reports and peer reviewed publications on the following range of in-vitro pyrogen tests:

1. Human Whole Blood IL-1,
2. Human Whole Blood IL-6,
3. PBMC IL-6,
4. MM6 IL-6, and
5. Human Cryopreserved Whole Blood IL-1,

it is concluded that these tests have been scientifically validated for the detection of pyrogenicity mediated by Gram-negative endotoxins, and quantification of this pyrogen, in materials currently evaluated and characterized by rabbit pyrogen tests.

These methods have the potential to satisfy regulatory requirements for the detection and quantification of these pyrogens in these materials subject to product-specific validation.

The test methods have the capacity of detecting pyrogenicity produced by a wider range of pyrogens, but the evidence compiled for, and considered within this peer review and validation process, is not sufficient to state that full scientific validation of this wider domain of applicability has been demonstrated and confirmed.

Thus, the above test methods can currently be considered as full replacements for the evaluation of materials or products where the objective is to identify and evaluate pyrogenicity produced by Gram-negative endotoxins, but not for other pyrogens.

This endorsement takes account of the dossiers prepared for peer review; the views of independent experts who evaluated the dossiers against defined validation criteria; supplementary submissions made by the Management Team; and the considered view of the Peer Review Panel appointed to oversee the process.

Thomas Hartung
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra

21 March 2006

1. The ESAC was established by the European Commission, and is composed of nominees from the EU Members States, industry, academia and animal welfare, together with representatives of the relevant Commission services.

   This statement was endorsed by the following Members of the ESAC:

   Prof Helmut Tritthart (Austria)
   Dr Dagmar Jírová (Czech Republic)
   Prof Elisabeth Knudsen (Denmark)
   Dr Timo Ylikomi (Finland)
   Prof André Guillouzo (France)
   Dr Manfred Liebsch (Germany)
   Dr Efstathios Nikolaidis (Greece)
   Dr Katalin Horvath (Hungary)
   Prof Michael Ryan (Ireland)
   Dr Annalaura Stammati (Italy)
   Dr Mykolas Maurica (Lithuania)
   Prof Eric Tschirhart (Luxembourg)
   Dr Jan van der Valk (The Netherlands)
   Dr Dariusz Sladowski (Poland)
   Prof Milan Pogačnik (Slovenia)
   Dr Argelia Castaño (Spain)
   Dr Patric Amcoff (Sweden)
   Dr Jon Richmond (UK)
   Dr Odile de Silva (COLIPA)
   Dr Julia Fentem (ECETOC)
   Dr Nathalie Alépée (EFPIA)
   Prof Robert Combes (ESTIV)
   Dr Maggy Jennings (Eurogroup for Animal Welfare)
   Mr Roman Kolar (Eurogroup for Animal Welfare)

   The following Commission Services and Observer Organisations were involved in the consultation process, but not in the endorsement process itself.

   Mr Thomas Hartung (ECVAM; chairman)
   Mr Jens Linge (ECVAM; ESAC secretary)
   Mr Juan Riego Sintes (ECB)
   Ms Beatrice Lucaroni (DG Research, Unit F.5)
   Mr Sylvain Bintein (DG Environment, Unit C.3)
   Mr Sigfried Breier (DG Enterprise, Unit F.3)
   Prof Dr Constantin Mircioiu (Romania)
   Dr William Stokes (NICEATM, USA)
   Prof Dr Vera Rogiers (ECOPA)

Annex

The novel pyrogen tests are based on the human fever reaction. Monocytoid cells, either primary from human blood or as propagated cell lines, detect pyrogens of different chemical nature and respond by the release of inflammatory mediators such as cytokines. Since lipopolysaccharides from Gram-negative bacteria are the only type of proven pyrogen, for which an International reference material is available, the tests were standardised to detect the presence of significantly less than 0.5 Endotoxin Units of this preparation, which is considered to be the threshold level for fever induction in the most sensitive rabbit species according to pharmacopoeia test procedures.
The five tests which were sufficiently reproducible and exceeded the rabbit test with regard to sensitivity and specificity for the detection of lipopolysaccharide spiked samples, differ with regard to cell source and preparation, cryopreservation and cytokine measured. The tests have been described elsewhere (1-4). The concept of the validation study (5) and the international validation studies are available (6-7).

1. Poole, S., Thorpe, R., Meager, A., Hubbard, A.R., Gearing, A.J. (1988) Detection of pyrogen by cytokine release. Lancet 8577, 130.

2. Taktak, Y.S., Selkirk, S., Bristow, A.F., Carpenter, A., Ball, C., Rafferty, B., Poole, S. (1991) Assay of pyrogens by interleukin-6 release from monocytic cell lines. J. Pharm. Pharmacol. 43, 578.

3. Hartung, T., Wendel, A. (1996) Detection of pyrogens using human whole blood. In Vitro Toxicol. 9, 353.

4. Schindler S, Asmus S, von Aulock S, Wendel A, Hartung T and Fennrich S. (2004) Cryopreservation of human whole blood for pyrogenicity testing. J. Immunol. Meth. 294, 89-100.

5. Hartung, T., Aaberge, I., Berthold, S., Carlin, G., Charton, E., Coecke, S., Fennrich, S., Fischer, M., Gommer, M., Halder, M., Haslov, K., Jahnke, M., Montag-Lessing, T., Poole, S., Schechtman, L., Wendel, A., Werner-Felmayer, G. (2001) Novel pyrogen tests based on the human fever reaction. The report and recommendations of ECVAM Workshop 43. European Centre for the Validation of Alternative Methods. Altern. Lab. Anim. 29, 99.

6. Hoffmann S, Peterbauer A, Schindler S, Fennrich S, Poole S. Mistry Y, Montag-Lessing T, Spreitzer I, Loschner B, vam Aalderen M, Bos R, Gommer M, Nibbeling R, Werner-Felmayer G, Loitzl P, Jungi T, Brcic M, Brugger P, Frey E, Bowe G, Casado J, Coecke S, de Lange J, Mogster B, Naess LM, Aaberge IS, Wendel A and Hartung T. (2005) International validation of novel pyrogen tests based on the human fever reaction. J. Immunol. Meth. 298, 161-173.

7. Schindler S, Spreitzer I, Loschner, Hoffmann S, Hennes K, Halder M, Brügger P, Frey E, Hartung T and Montag T. (2006) International validation of pyrogen tests based on cryopreserved human primary blood cells. J. Immunol. Meth. 316, 42-51.

[This Page Intentionally Left Blank]

Appendix C2

Press Release: "Fewer Tests on Animals and Safer Drugs: New EU Tests Save 200,000 Rabbits per Year"

[This Page Intentionally Left Blank]

IP

Brussels, 12 May 2003

Fewer tests on animals and safer drugs: new EU tests save 200,000 rabbits per year

New, groundbreaking methods of drug testing to replace animals with safe alternatives, saving up to 200,000 rabbits per year, were unveiled today in Brussels by European Research Commissioner Philippe Busquin. The set of six tests detects potential fever-causing agents (pyrogens) in drugs, by using human blood cells instead of rabbits. The new tests have been developed by a EU-supported research team, involving national control laboratories, test developers, and companies. The tests are being validated by the Commission. They are already being used in over 200 laboratories across the world. Thanks to these alternative methods rabbits will no longer be needed to test the presence of pyrogens in parenteral (non oral) drugs.

"The use of animals to test drugs is unfortunately necessary to safeguard human health," said European Research Commissioner Philippe Busquin. "But we can reduce, replace and refine animal testing, with EU-sponsored research leading the way at world level. The EU's validation of these new testing methods will encourage their broad take-up by industry, ensure drug safety and quality, and reduce the use of animal research. This is an example of the European Research Area in action, developing an environment in which scientific results can be rapidly exploited and transformed into products and processes that improve quality of life, increase competitiveness and benefit animal welfare."

The safety and potency of commercially available medicines and vaccines must be guaranteed. Innovative research, funded and validated by the Commission, aims to replace existing animal-based test methods for fever-causing agents (pyrogens) in parenteral drugs with a new generation of in vitro tests that are more accurate, quicker and more cost-effective.

Blood cells replace rabbits

Understanding of human immunology has advanced rapidly in the past 20 years. Work on human fever reaction and development of test systems for fever mediator molecules, combined with improved cell biology techniques, now enables the innovative use of human cells as biosensors for pyrogens (fever-causing agents). The EU study[1] set out to compare and harmonise six in vitro assays to develop a "state-of-the-art" method for inclusion into the European Pharmacopoeia - which sets the requirements for the quality control of drugs in Europe - thus improving consumer safety.

The EU role

[1] Cell factory project: *Comparison and validation of novel pyrogen tests based on the human fever reaction, with a view to the ultimate replacement of the rabbit pyrogen test and the Limulus assay* (QLK3-1999-00811)

The research project funded by the Commission under the EU Fifth Research Framework Programme (1998-2002) brought together the best teams from academia, industry and regulatory bodies. The Commission's Joint Research Centre (the "ECVAM" facility, or "European Centre for Validation of Alternative Methods") played a major role in the project through provision of scientific and technical advice on the design of the validation study, application of good laboratory practice procedures and distribution and coding of test material.

Industry and regulators jump on board

Interest from both regulatory authorities and industry is very high, with many contributions coming from outside the project consortium that included national control laboratories, test developers, a major pharmaceutical company and a producer of diagnostic kits. For example, the European Pharmacopoeia has set up an international expert group to draft a general method on these new tests. In fact, the tests are already in use in about 200 laboratories worldwide, with great success.

Further take-up and new applications

The Commission will take responsibility for further application of this multidisciplinary, international validation study, including an intended patent. This will encourage successful transfer of the tests and help open new fields for pyrogen testing, such as cellular therapies, medical devices and pollution control in the work place.

Reducing, replacing or refining animal experimentation

Drug quality control is a trans-national matter, which is standardised and regulated in Europe at EU level, thus requiring international collaborative efforts. The European Commission ensures full support for applications to reduce, replace or refine animal experimentation as required by the 1986 Council Directive[2]. This aim is echoed by the European Pharmacopoeia. The "Three Rs" provide a strategy to minimise animal use, without compromising the quality of the scientific work being done.

ECVAM's role is to co-ordinate international validation studies, act as a focal point for the exchange of information, to set up and maintain a database on alternative methods, and to promote dialogue among legislators.

Background: pyrogen and non-oral drugs

Parenteral drugs are commonly employed throughout Europe for treating a variety of illnesses. Ensuring the safety of such widely used drugs requires strict monitoring and control against any possible pyrogenic contamination on a batch-by-batch basis. The most important pyrogen is endotoxin, a constituent of the cell wall of gram-negative bacteria that can generate endogenous fever mediators by white blood cells, particularly monocytes and macrophages.

Rabbits or…

In the rabbit pyrogen test, the test substance is injected into rabbits and any subsequent change in body temperature recorded. A significant rise in temperature indicates the presence of pyrogens. While it has served drug safety control for more than 50 years, it fails for important new therapies such as cellular products or species-specific agents.

[2] *Novel in-vitro testing as alternatives to animal testing*; Council Directive 86/609/EEC

… horseshoe crabs?

Until now, the only in vitro alternative available is the LAL test, based on coagulation of blood from the horseshoe crab (Limulus polyphemus). However the LAL test detects only one class of pyrogens – endotoxins from gram-negative bacteria – leaving patients at risk from "non-endotoxin" pyrogens such as gram-positive toxins, viruses and fungi. It is also subject to interference by various non-pyrogenic substances. And, as it is based on the defence system of an arthropod, it cannot provide results perfectly relevant to humans.

No – human blood cells!

Six alternative cellular assays have therefore been developed to replace the animal rabbit pyrogen test and close the safety gap presented by use of the LAL test in controlling parenterals. All these test systems are based upon the response of human leukocytes (principally monocytes), which release inflammatory mediators (endogenous pyrogens) in response to pyrogenic contamination (exogenous pyrogens).

Quicker, more accurate and more effective

The new tests have several advantages compared with the rabbit test: they are less laborious, cheaper and more sensitive. Results of the validation study suggest that testing on animals can be completely replaced. In contrast to the LAL, the new assays are not restricted to endotoxins from gram-negative bacteria but detect all classes of pyrogens and reflect the potency of different endotoxins in mammals, without suffering interference from endotoxin-binding components in blood products. A commercial kit version for one of the assays has already been developed and standardised, and pre-tested cryopreserved (frozen) blood as a versatile test reagent containing the blood cells as biosensors is under development.

For further information please visit:

http://ecvam.jrc.it/index.htm

http://europa.eu.int/comm/research/quality-of-life/cell-factory/volume1/projects/qlk3-1999-00811_en.html

[This Page Intentionally Left Blank]

Appendix C3

ECVAM Replies to Questions of ICCVAM Pyrogenicity Peer Review Panel

[This Page Intentionally Left Blank]

1 February 2007

ECVAM replies to questions of ICCVAM Pyrogenicity Peer Review Panel

1. Availability of ESAC Peer Review Report

Since we are creating a precedent in making ESAC peer-reviews public, a discussion within ESAC is required, especially since a number of external experts have been involved, who have not been asked. Thus, we are unfortunately not able to make this available at this stage of the process.

2. Lot numbers

a) e-mail of David Allen on 10/01/2007
replied on 12/01/2007 with list of drugs as PDF attached

VALIDATION STUDY: LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	21402
MCP Hexal	Hexal	21JX22
Orasthin	Hoechst	W015
Sostril	Glaxo Wellcome	1L585B
Traubenzuckerlösung 5% Eifelfango	Eifelfango	1162

Alkohol-Konzentrat 95% = aethanol 95%
Traubenzuxkerloesung 5% Eifelfango = 5% glucose solution

b) e-mail of David Allen on 12/01/2007
Yes, individual lots were tested in all methods during the validation/catch-up validation study

However, some of the lots used in the validation study were no longer available for the catch-up validation study and one product (Orasthin) was no longer on the market. It was replaced with a product (Syntocinon) containing the same active ingredient. Please find attached the pdf file "List of drugs catch-up validation" and the table below highlighting differences in lot numbers and products.

VALIDATION STUDY (CATCH UP): LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	26803
MCP Hexal	Hexal	21JX22
Orasthin	Hoechst	not available
Sostril	Glaxo Wellcome	3H01N
Syntocinon 3 I.E.	Novartis	S00400
Traubenzuckerlösung 5% Eifelfango	Eifelfango	3132

Alkohol-Konzentrat 95% = aethanol 95%
Traubenzuxkerloesung 5% Eifelfango = 5% glucose solution

3. GLP concordance

e-mails of David Allen on 9/01/2007 and on 12/01/2007 (question 1)

a) In vitro data

The initial validation study has been carried out to large extent in laboratories such as National Control laboratories, which do not operate under GLP. It was, however, agreed to comply with the requirements of GLP, especially with regard to the creation and management of SOPs. The partner laboratories have received presentations on the requirements. No auditing was done but various quality checks and blinding mainly under the responsibility of ECVAM were included.
In the catch-up validation, two GLP laboratories and two National Control laboratories participated.

Raw data: In both studies the laboratories were asked to transfer the readings into the excel sheets provided by the biostatistician. This was mostly done by directly inserting the ASCII files created by the plate reader. However, reader printouts are available and can be provided on request.

b) In vivo data and reference to Section 4.4 in the ECVAM BRDs
Indeed it should read here "not applicable" as stated in the WB/IL-6 BRD, since the RBT was not performed during the validation study. As indicated in 4.1 the data used were provided by the Paul-Ehrlich-Institute (PEI; www.pei.de), which is the German Federal Agency for Sera and Vaccines (competent authority) and conducts the RBT according to the European Pharmacopoeia. For further information on the quality assurance established at the PEI please contact Dr Thomas Montag (e-mail: month@pei.de).

This should also be corrected in the main document 4.4 In vivo data quality.

4. Data analysis
e-mail of David Allen 12/01/2007 – question 2

The same data analysis was applied in both studies. The first paragraph in Section 5.3 reads *A generally applicable analytical procedure was employed. This procedure includes a universal PM as well as quality criteria. First, a two-step procedure consisting of a variance-criterion and an outlier-test was applied. For this, the Dixon's test (Barnett and Lewis, 1984), which is USP approved, was chosen with the significance level of* α*=0.01 and applied to identify and eliminate aberrant data.*
Please find attached to our mail, the document 'Trial data report' of the validation study. It was not included in the submission to ICCVAM, since a lot of the procedures described here are included in the BRD. Related to your question, you will find in Chapter 4.2 the procedure describing the exclusion of data. There, the Coefficient of Variation (CV) was used as a trigger to investigate the replicates of a given control or sample. Excessive variability would severely impair the prediction model, resulting mainly in a loss of specificity. The CVs were empirically determined for each assay based on the information collected in the protocol optimization phases (Phase A & B) and the prevalidation. Thus, they can differ between assays.
In addition, we attach the document *Analytical procedure to identify and eliminate outlying observations* written by the responsible statistician, Sebastian Hoffmann, during the validation study and which gives rationales for applying this procedure.

5. Selection of test substances
e-mail of David Allen 12/01/2007 – question 3

Please find attached the file "Rationale for selection of test substances".

6. Removal of DMSO
e-mail of David Allen 12/01/2007 – question 4

Schindler et al 2004 state:

We sought to develop a protocol which would allow the use of the thawed whole blood samples directly without any washing steps to remove the cryoprotectant, as such a step would eliminate the essential advantages of the human whole blood assay, i.e., the ease of performance which allows a high degree of standardization as shown for various applications (Fennrich et al., 1999). Furthermore, besides stress and handling artifacts, the cells would lose their autologous plasma that permits a number of physiological responses, e.g., the sensitive response to

lipopolysaccharides (endotoxin, LPS) via lipopolysaccharide binding protein (LBP; Schumann, 1992; Fenton and Golenbock, 1998).

Indeed DMSO is not removed and up to now artefacts attributed to the presence of DMSO were not observed. The presence of DMSO enhances the IL-1 production and leads to a delay in the release. The fact that no wash step is required reduces strongly variation and introduction of artefacts.

7. Possible cytotoxicity
e-mail of David Allen 12/01/2007 – question 5

The aspect of cytotoxicity is covered by interference testing. As stated already on various occasions, interference testing (what we called positive product control in the validation study) is a must before you can use the WB (and the other) assays. If a substance would interfere with the assay by being cytotoxic, the spike recovery would be below 50%.

8. Freeze-thaw step for CRYO WB/IL-1 (Konstanz method)
e-mail of David Allen 12/01/2007 – question 6

This question was already posed during the drafting of the ICCVAM peer review documents (Mail David Allen 1/08/2006 question 3 and attached document PyroProtocol31Jul06)

In our reply (sent on 8/09/2006 with attachment reply_PryoProtocol31Jul06), we stated on page 2: *The freezing thawing enhances the IL-1 release and makes the Konstanz method more robust and reliable. It is not needed for the PEI method since the IL-1 release levels are higher.*
In fact, it should read that the freezing thawing enhances the IL-1 yield since the IL-1 produced in but not released by the monocytes is also measured.
It has been shown by Boneberg and Hartung (2003) that 10fold higher concentrations of (pro-)IL-1ß are found when including intracellular cytokine by whole blood lysis:

Ref *Boneberg E. and Hartung T. Febrile temperatures attenuate IL-1β release by inhibiting proteolytic processing of the proform and influence Th1/Th2 balance favoring Th2 cytokines. J. Immunol. 2003, 171:664-8.* attached.

Appendix C4

Rationale for the Selection of the 10 Substances Tested in the Validation/Catch-Up Validation Study of *In Vitro* Assays for Pyrogen Testing

[This Page Intentionally Left Blank]

Rationale for selection of the 10 substances tested in the validation/catch-up validation study of in vitro assays for pyrogenicity testing

Selection committee:
Thomas Montag-Lessing (chair), Michael Jahnke, Ingeborg Aarberge, Sandra Coecke

The main points which led to the selection were stability of the spikes, relevance, availability and costs of the substances:

1. Stability of the spikes, coding, interference testing

Experiments to evaluate the stability of endotoxin spikes in the final products revealed that stability of low endotoxin concentration could not be guaranteed over the time period needed for the prevalidation/validation study. Therefore, endotoxin spikes in higher (stable) concentrations were produced, filled in separate vials and coded. The laboratories received the clean substance plus the coded spikes, the clean substance had to be used for interference testing and contaminated with the coded spikes for the actual tests.

2. Relevance

The absence of pyrogens is crucial for intravenously administered drugs, this is reflected in the rabbit pyrogen test where the test substance is injected into the ear vene.
Therefore, only substances intended for i.v. injection were selected. In addition, it could be evaluated whether the in vitro assays would be able to detect 0.5 IU/ml endotoxin, which corresponds to threshold inducing fever in rabbits.

3. Availability/feasibility

- Substances should be on the market, thus the final product in the original vials could be tested and the conditions under which a lab performing final lot release would work could be met, e.g.
    - avoid possible contamination with pyrogens during opening the vials, drawing the samples etc
    - performing interference testing (corresponds to positive product testing in the validation trial)

- One lyophilised product was included in order to check for potential failures (e.g. pyrogenic contamination during reconstitution of the drug)

- Substance not interfering with any of the assays in order to control the correctness of the spiking procedure. Therefore, 0.9% NaCl pyrogen-free solution was included (Drug A and B)

4. Costs

- Due to the restricted funds available, costs of the substances to be tested in the validation trial played a role, e.g. it was not possible to include a rather expensive blood product as coagulation Factor VIII (Haemate® was used in the prevalidation trial).

Appendix C5

Comparison and Validation of Novel Pyrogen Tests Based on the Human Fever Reaction: Trial Data Report

[This Page Intentionally Left Blank]

Validation of Biomedical Testing Methods

Comparison And Validation Of Novel Pyrogen Tests Based On The Human Fever Reaction

Acronym: Human (e) Pyrogen Test

Trial data report

Date of Circulation:
File Identifier: *Stp-HPTVv03*

JOINT RESEARCH CENTRE
EUROPEAN COMMISSION
Institute for Health and Consumer Protection
ECVAM

TABLE OF CONTENTS Page no.

Comparison And Validation Of Novel Pyrogen Tests Based On The Human Fever Reaction

Acronym: Human (e) Pyrogen Test

1 INTRODUCTION

The objective of the "Human(e) Pyrogen Test" project is to assess the performance and use of six recently developed *in vitro* pyrogen tests. These tests are based on the human fever reaction. As they are meant to be similar to the currently used Rabbit Test, the analytical procedure is designed to give a dichotomous outcome. In detail, drugs, which have to be tested for pyrogenic contamination due to regulatory instruction, have to be classified either as hazardous, i.e. pyrogenic, or as safe for humans. Hence, securing the safety of humans is the primary objective of pyrogenicity testing in general. Therefore, the prediction model is constructed to give a clear-cut classification of a given drug taking the safety aspect into account.

In addition to the prediction model, procedures to ensure quality criteria the test systems have to meet are included. A two-step method to identify and eliminate aberrant data as well as a test for a sufficient limit of detection are provided. Information from previous phases of the project established the basis to develop and define these methods. The data from the pre-validation study were used to refine the procedures.

All methods of the analytical procedure were developed to be applicable to each of the six test systems and were accepted by the participants.

.

2 THE BASIC BIOLOGICAL PRINCIPLE

The six test systems make use of the same biological principle. The mediators of the human fever reaction are cytokines, which are produced by monocytes. This principle is employed by incubating either fresh human cells or cell lines with the drug to be analysed under SOP-defined conditions. As there are several cytokines, which highly correlate with the human fever reaction, the cytokine of choice of the test sytems differs. Mainly the cytokine IL-6, but also IL-1β, TNF and neopterin were chosen as endpoints. After the incubation, an also SOP-defined ELISA-step is performed. In this step the cytokine is bound, visualised and finally measured by an optical reader. As the visualisation of the endpoint, measured as optical density (OD), is proportional to the amount of cytokine present, the resulting data are metrically scaled. In these entire procedure, a biological standard, WHO-LPS 94/580, is employed as an objective tool for comparison.

One of the test systems is based on a competitive ELISA, which results in a monotonically decreasing dose-response relationship, whereas the other systems show an increasing relationship due to their sandwich ELISA technique.

3 STATISTICAL PROPERTIES OF THE DATA

In previous phases the statistical properties of the data were analysed. Replicate observation for a fixed known control or an unknown drug revealed a right-skewed distribution. In experiments with large sample sizes it was shown, that a ln-transformation of the raw OD-data allows to assume a gaussian distribution of the data, which parameters can be estimated by the mean and the empirical variance.

As handling errors in the conduction of the test result in extreme observations, which may have an crucial impact on the prediction model, the probability of occurrence and impact of these observation was analysed. Although the probability of extreme observations is small for all tests, the inclusion of a method to identify and eliminate these data is indicated to ensure an optimised performance of the prediction model.

Furthermore, it was confirmed, that the dose-response relation ship between concentration of the contamination and the response increases, respectively decreases, monotonically for increasing concentration.

4 METHODS

4.1 Background

The entire analytical procedure consists of three different techniques, two of which assure the appropriateness of the data. The ELISA-plates employed have a 96-well format. The data of one plate have to be considered as a whole, which can not be compared to other plates due to uncontrollable variation. Therefore, each plate has to include all controls required for the analytical procedure. These are a negative control, which is 0.9%-NaCl, and a positive control of the WHO-LPS 94/580 standard diluted in 0.9%-NaCl, as well as negative and positive controls of the drugs, which are to be tested on the plate. Negative controls of a drug are obtained by released batches of the drugs. Positive controls are gained by adding 0.5 endotoxin units (EU)/ml of WHO-LPS 94/580. These 0.5 EU/ml were concordantly defined as the threshold concentration of endotoxin that induces fever in humans under worst conditions. In previous experiments it was shown, that this positive control lies in the most sensitive region, i.e. the steepest part, of the dose-response curve of all six test systems. In the following the NaCl-controls are denoted as "C-" (negative) and "C+" (positive). Similarly, the controls of a drug S_i are denoted as "S_i-" ans "S_i+". Furthermore, "S_ij", j=1, 2, 3, 4, 5, represent the blinded versions of the drug S_i.

4.2 Method A: Identification and elimination of aberrant data

The first method to be applied is an method to check the quality of the data of a plate. In general, this is done by a two-step procedure, which firstly identifies the sets of replicates with an extremely large variation. A set of replicates consists of four replicates per control, respectively drug tested. For every test system a maximal coefficient of variation (CV_{max}) was extracted from the available information. If the CV of a set of replicates is smaller than its CV_{max}, it is analysed as it is. Otherwise, the set is examined in the second step. This second step is a test for outliers. Therefore, the Dixon's test (1), which is USP approved, was chosen with the significance level of α=0.01. Preliminary to the testing itself, the raw OD-data are transformed with the natural logarithm, which normalises the data to meet the

prerequisites of the Dixon's test. If one observation in a set, which is identified by the Dixon's test, is responsible for its large variation, then this observation is excluded. If the variation is due to all observations, i.e. the absence of an outlier, the entire set of replicates is excluded from further analysis. Unfortunately, this approach poses the danger, that a whole plate can not be analysed, when a control is to be excluded. Therefore, both steps were chosen conservatively. Nevertheless, the empirical nature of the first step is not optimal and depends on general properties of the test system. But all established statistical methods, which address this problem, e.g. the Bartlett test for heterogeneity of variances, are not appropriate, because the variance structure over the range of concentration is highly variable and their global character. In table 1 the empirically derived CV_{max} are listed for the six test systems. The approach could be harmonized over all test systems.

test system	MM6	PBMC	THP-1BN	THP-1IK	WBT-KN	WBT-NI
CV_{max}	0.25	0.45	0.45	0.25	0.45	0.45

Table 1: Maximum CV's for the six test systems

4.3 Method B: Assuring the limit of detection

The second method is designed to ensure an minimum limit of detection of a plate (2). Because of the pre-defined dichotomous classification, a crude criterion, which merely shows strict monotonicity in the interesting part of the dose-response curve, can be chosen. Therefore, a one-sided t-test with a significance level of α=0.01 is applied to the ln-transformed data to ensure, that the response to the positive control is significantly larger than that of the respective negative control.

4.4 Method C: The prediction model

The third and most important statistical tool is the so-called prediction model (PM). In general, it is a statistical model, which classifies a given drug by an objective diagnostic or deciding rule. The objective of a dichotomous result requires a clear cut PM, which assigns a drug in one of the two classes "pyrogenic for humans" and "non-pyrogenic for humans". As the members of the project decided on a threshold positive

control, a one-sided test is appropriate for the task. Because the data are normalised by a ln-transformation, a t-test was chosen. Although the variances over the range of concentration converge by the transformation, the assumptions of equal variances does generally not hold true, because it depends on additional covariates. Therefore, the one sided Welch-t-test (3) is applied. Due to the safety aspect of the basic problem, the hypotheses of the test are

$$H_0 : \mu_{S_i j} > \mu_{S+} \quad vs \quad H_1 : \mu_{S_i j} < \mu_{S+},$$

where $\mu_{...}$ denotes the parameter of location of the respective ln-transformed distribution. This approach controls the probability of false positive outcomes directly by means of its significance level α, which is chosen as 0.01, because is assumes hazard, respectively pyrogenicity, of the tested drug in H_0, and assures safety, i.e. non-pyrogenicity. The test statistic is

$$T_{S_i j} = \frac{\bar{x}_{S+} - \bar{x}_{S_i j}}{\sqrt{\frac{s^2_{S+}}{n_{S+}} + \frac{s^2_{S_i j}}{n_{S_i j}}}}.$$

The PM is built by means of the outcome of the test. Let 0 denote safety and 1 denote hazard. The classification of S_i-j is then determined by

$$S_i j = 0, \text{ if } \quad T_{S_i j} > t_{0.99; n_{S+} + n_{S_i j} - 2},$$

$$S_i j = 1, \text{ else,}$$

where $t_{0.99; n_{S+} + n_{S_i j} - 2}$ the 0.99-quantile of the t-distribution with $n_{S+} + n_{S_i j} - 2$ degrees of freedom. The number of replicates for every control and sample, i.e. $n_{...}$, was harmonised for all test systems to be four. Due to the possibility of removing one observation by the outlier test, the number of replicates could be reduced to three. The classification of a version of a drug is regarded as an independent decision. Therefore, the niveau α is local.

4.5 Method D: 2x2 contingency tables for the final results

Finally, the classifications of the drugs will be summarised in 2x2 contingency tables, formally presented in table 2.

		pre-defined class		Σ
		1	**0**	
classification by test system and PM	**1**	**a**	**b**	**a+b = $n_{.1}$**
	0	**c**	**d**	**c+d = $n_{.0}$**
Σ		**a+c = $n_{1.}$**	**b+d = $n_{0.}$**	**n**

Table 2: 2x2 contingency table

From these tables estimates of the sensitivity, i.e. the probability of correctly classified positive drugs, and specificity, i.e. the probability of correctly classified negative drugs, will be obtained by the respective proportions. Furthermore, these estimates will be accompanied by confidence intervals, which will be calculated by the Pearson-Clopper method (4). For example, let $\hat{p}_{SE}$ denote the proportion, namely the sensitivity, under investigation. Then the confidence interval to a niveau α is calculated as

$$\left[p_{SE}^{L} = \frac{\mathrm{a} F_{2\mathrm{a};2(n_{1.}-\mathrm{a}+1);\frac{\alpha}{2}}}{n_{1.}-\mathrm{a}+1+\mathrm{a}F_{2\mathrm{a};2(n_{1.}-\mathrm{a}+1);\frac{\alpha}{2}}}; p_{SE}^{U} = \frac{(\mathrm{a}+1)F_{2(\mathrm{a}+1);2(n_{1.}-\mathrm{a});1-\frac{\alpha}{2}}}{n_{1.}-\mathrm{a}+(\mathrm{a}+1)F_{2(\mathrm{a}+1);2(n_{1.}-\mathrm{a});1-\frac{\alpha}{2}}} \right],$$

where $F_{\ldots}$ denotes the respective quantile of the F-distribution and $n_{1.}$ is the sample size of the positive drugs and a the number of correctly classified drugs. By contaminating the drugs artificially and by defining a threshold dose, which is assumed to be appropriate, the class of a drug is determined beforehand. The versions of drugs, which are effectively contaminated, but below the threshold dose, are considered to be negative, respectively safe, because their contamination is not crucial for humans in terms of endotoxin limit concentration.

5 ANALYTICAL PROCEDURE

5.1 General procedure

The process of the analytical procedure is highlighted in figure 1. Firstly, the data of the controls of a ELISA-plate are checked for aberrant data with procedure A. If indicated, outliers are removed. If sets of replicates are to be removed, this is recorded, but due to the empirical base of the first step of A, the data will be further analysed with reservations. Afterwards, the remaining data of the controls are tested with method B to ensure a minimum limit of detection. If the controls in 0.9%-NaCl do not differ significantly, the further analysis is done with reservations. If the controls of a drug do not differ significantly, all data of this drug do not qualify for further analysis. The last part of quality assurance is the application of method A to the data of the blinded drugs. Here, drugs, which fail the criteria, are removed from further analysis. Finally, the remaining data are put to the prediction model. The classification of the still blinded drugs are sent to ECVAM in an official document, which will in general comprise the assigned class for every drug structured by test system, laboratory and drug. Upon receipt ECVAM will send the blinding code in a electronically generated document by e-mail to the project's statistician Sebastian Hoffmann (e-mail: sebastian.hoffmann@uni-konstanz.de). Additionally, a hardcopy of the blinding code will be sent by post. Once the data are unblinded, the final results, which core will be method D, can be summarised, explicitly analysed and appropriately presented. Additionally to the contingency tables and related topics, an inter-laboratory comparison will be done. Furthermore, the reasons for misclassifications will be identified.

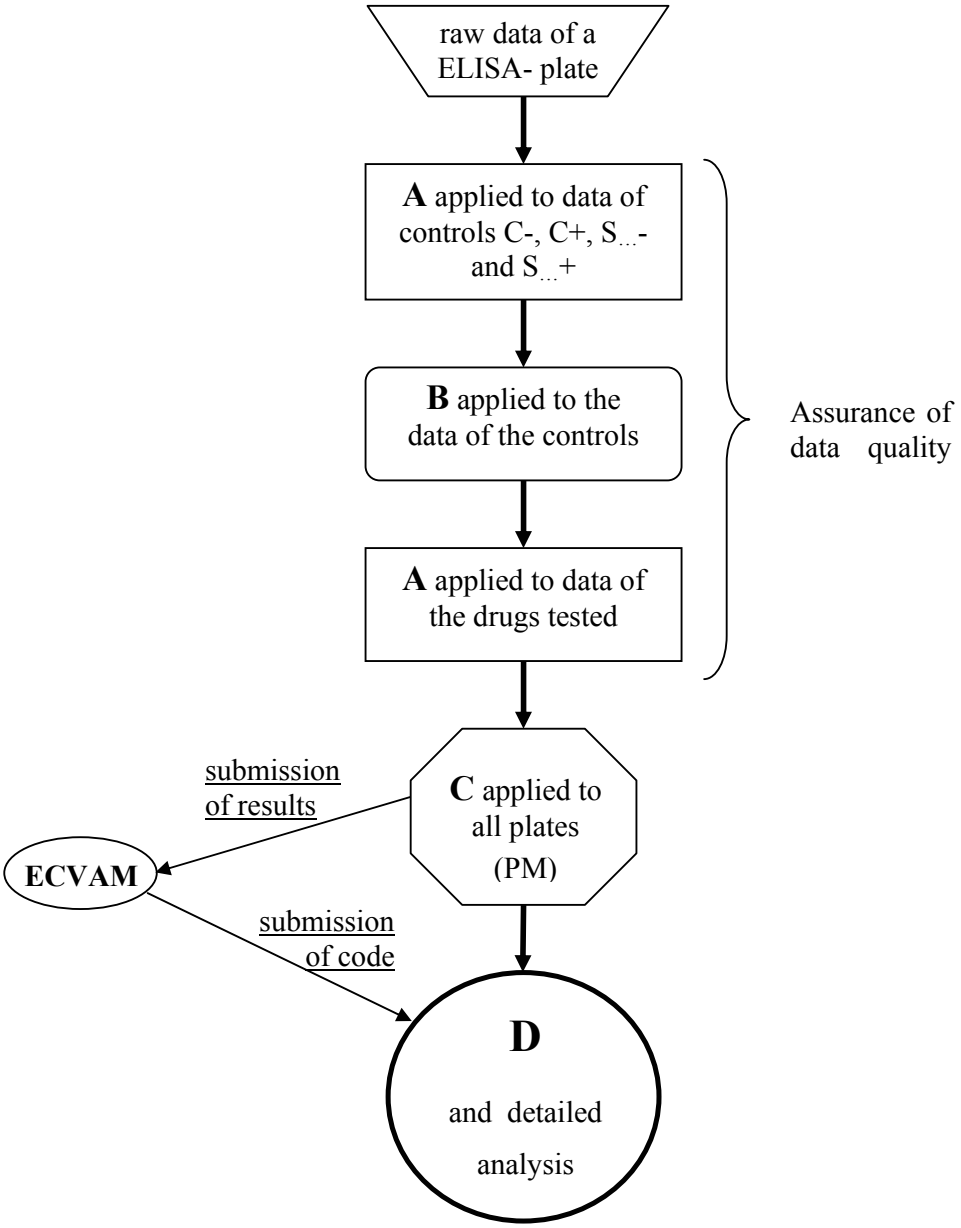

Figure 1: Flowchart of the main analytical procedure

5.2 Modification for the test systems PBMC and WBT-NI

This analytical procedure has to be modified for the two test systems PBMC and WBT-NI. The two test systems base their classification of a drug not on one outcome, but they classify a drug by the results of several independent experiments. Because both methods rely on fresh blood, the inter-donor variability is taken into account by using the blood of several donors and conducting the test independently. The modifications, which have to be made, arise out of contradictory classification of a drug by different donors. Therefore the classification of a drug is determined by the combination of the single donor-dependent results, which are calculated with the analytical procedure presented.

6 ADDITIONAL ANALYSIS

The detailed analysis will contain an inter-lab comparison per test system, whereas measures of correlation as well as similarity can be used. Furthermore, shortcomings depending on specific drugs, laboratories, the analytical procedure and/or test systems will be examined by exploratory statistical methods.

Additionally, modifications of the methods A, B and C will be used to optimise the analytical procedure with the information from the new data. In general, these are the consequences of more restrictive or less restrictive assumptions. With regard to A, the results of a procedure without a tool for aberrant data will be compared to the results of the described procedure. For B, more restrictive criteria to ensure a valid dose-response relationship will be applied, e.g. techniques for ratios between controls based on Fieller's theorem (5, 6). Besides, modifications in the t-test of the prediction model will be of interest, mainly assumptions considering the variance and a multiple testing approach. E.g. a simulation, which allows for the $\sqrt{k}$-rule optimising the Dunnett's test could be realisable.

Finally, methods taking the real life situation of pyrogen testing into account will be highlighted. These include a Fieller-based method to handle interference.

7 REFERENCES

1. Barnett, V., Lewis, T. (1984). *Outliers in statistical data*. 2nd edition, Chichester, US: John Wiley & Sons, pp 171-172.
2. Hothorn, L. A. (1995). Biostatistical analysis of the 'control vs. k treatments' design including a positive control group. In *Testing principls in clinical and preclinical trials*, pp 19-26, Stuttgart, Germany: Gustav Fischer Verlag.
3. Snedecor, G. W., Cochran, W. G. (1989). *Statistical methods*, 8th edition, Ames, US: The Iowa State University Press.
4. Clopper, C. J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-413
5. Fieller E. (1954) Some problems in interval estimation. *J. Royal Statistical Society* **B16**, 175-185.
6. Hauschke, D., Hothorn, T., Schaefer, J. (2002). The role of control groups in mutagenicity studies: matching biological and statistical relevance. *ATLA* (in press)

Appendix C6

List of Drugs for the Catch-Up Validation Study

[This Page Intentionally Left Blank]

VALIDATION STUDY (CATCH UP): LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%[1]	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	26803[2]
MCP Hexal	Hexal	21JX22
Orasthin[3]	Hoechst	not available
Sostril	Glaxo Wellcome	3H01N[4]
Syntocinon 3 I.E.	Novartis	S0040
Traubenzuckerlösung 5% Eifelfango[5]	Eifelfango	3132[6]

16.02.2004

[1] "95% Alcohol Concentration"
[2] Different lot number
[3] Orasthin no longer on the market, replaced with Syntocin 3 I.E. containing also oxytocin
[4] Different lot number
[5] "5% Glucose Solution"
[6] Different lot number

[This Page Intentionally Left Blank]

Appendix C7

Analytical Procedure to Identify and Eliminate Outlying Observations

[This Page Intentionally Left Blank]

Analytical procedure to identify and eliminate outlying observations

Introduction

As seen in the pre-validation, the problem of outlying observations is not appropriately solved yet. The crucial issue about these observations is their impact on the prediction model, which could result in false classifications of substances. Of course, one could just neglect such observation, as done in the pre-validation. This is the most easy way, but its appropriateness with regard to sensitivity and specificity is depending on the probability of outlying observations. So far, this probability was low, smaller than 5%, for all assays. Nevertheless, even if there are only a few outliers, this approach has the disadvantage, that one would have to live the most extreme and maybe even pre-identified outlying observations, e.g. when a technician recognises that she/he made a gross handling error.

Therefore, a new analytical procedure was developed. First of all, the objective of such an procedure has to be defined precisely. On the one hand, a way to identify obvious handling errors, which most often can be identified by eye by trained persons, is needed. On the other hand, a method to handle sets of replicates, which are extremely untypical for a specific assay, has to be taken into account. Hence, a generally applicable two-step procedure is proposed.

Step 1: Checking the variation between sets of replicates

Firstly, the data of an ELISA-plate are checked for untypical variation of one or more sets of replicates. In the given situation, one is only interested in those sets with extremely large variance. In general, there are two situations which have to be considered. On the one hand, just one observation could be responsible for a huge variance in its set of replicates. On the other hand, equally distributed replicates over a large range of response, which includes the situation of two outliers in a set, might be the reason.

The existing statistical tests addressing this question, e.g. the Bartlett-test, are not appropriate due to various reasons, but mainly because they assume homogeneity of variances and show global heterogeneity. Thus a simple empirical method was derived for every assay, which is mainly based on the data of the pre-validation and the information from Phases A and B. The core of this method is an appropriate measure of variation for a set of replicates. Here, the CV is chosen, but the variance or the standard deviation can be used more or less equivalently. From these empirical information, a maximum CV, denoted as CV_{max}, was derived, which can be used as a tool to assess the variation of each set of replicates very easily. If a CV of a given set is larger than CV_{max}, then this set will be examined further in the second step of the

procedure. If the CV is smaller, then the data of the set will be analysed as they are. To highlight this empirical method, it is exemplarily explained for the Novartis-PBMC assay. In figure 1 the variation within sets of replicates for all available data is presented.

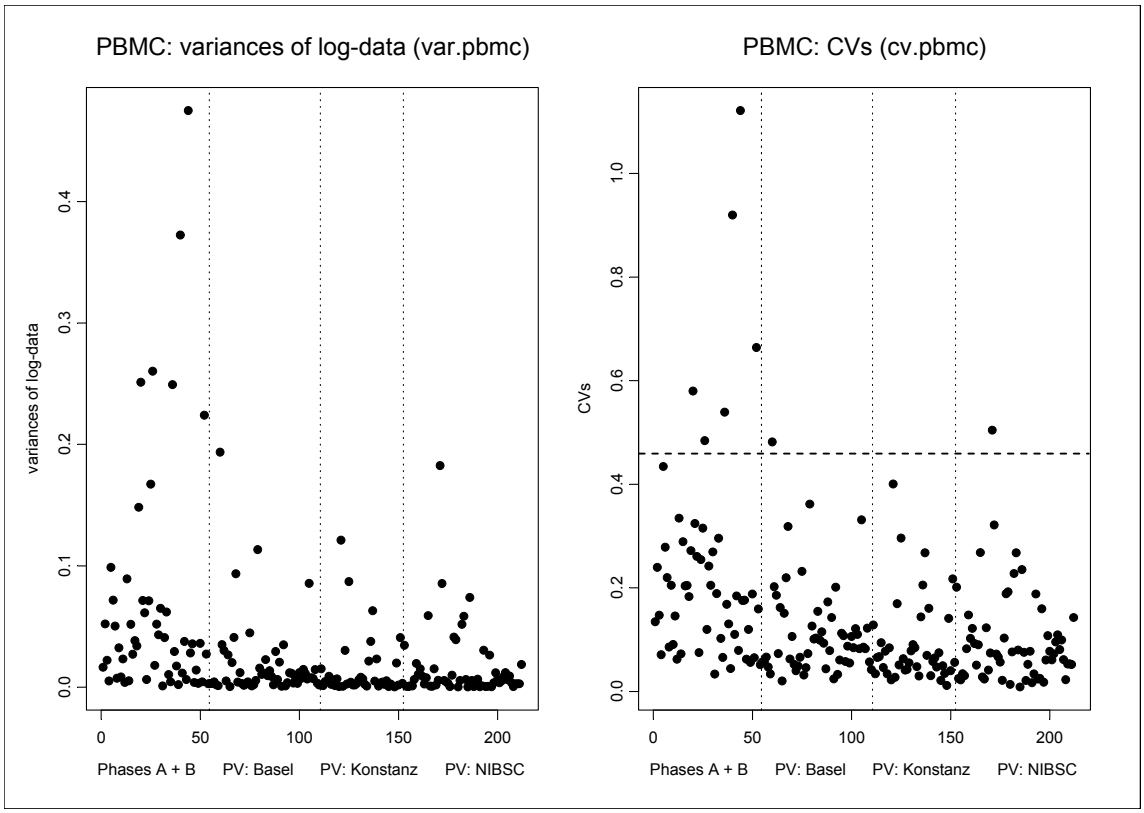

Figure 1: Variation within replicates for the PBMC-assay

Together with the raw data and some linear modelling techniques, here $CV_{max}(PBMC) = 0.45$ was chosen. This choice identifies two out of 154 pre-validation data sets as outlying sets in the right part of figure 1. As can be seen in the left part of figure 1, a criterion based on the variance, e.g. $Var_{max}(PBMC) = 0.18$ is almost identical, which even can be shown by some statistical approximation under certain assumptions.

Additionally to the approach with the CV, a criterion based on the ratio of variances was applied. Also having the empirical background, it did not show any advantage.

Step 2: Checking the variation within sets of replicates identified in step 1

Let S_i, i=1,…, denote the crucial, in step 1 identified sets with $CV(i) > CV_{max}$. In this second step the reasons for the high variation of the S_i's are examined. Firstly, a common test for outliers, the FDA-approved Dixon-test, is applied to each S_i with the niveau α of 1%. If an outlier is identified, it is withdrawn from its set and the remaining data are further analysed. If

no outlier is detected, the observations of a set of replicates are regularly distributed over a large range of response. In the latter case, it is recommended, to repeat the substance(s) S_i on another plate.

Discussion

Assuming such a partly empirical approach is appropriate, one still has to be aware of its properties and effects, especially when applied in the validation study. Firstly, the procedure gives excellent results when applied to the pre-validation data. This is expected, because the CV_{max} criterion was mainly derived by the data themselves, which makes it a self-fulfilling prophecy. Therefore, it poses the danger of choosing the CV_{max} too small, because it may lead to a lot of rejected sets. Additionally, maintaining such an empirical procedure demands to check regularly for the validity of the chosen $CV_{max.}$

Secondly, in the given situation of the validation study, the impact of the retrospectively applied procedure has to be taken into account. Because the data are checked outside the labs, the sample size of the number of classified samples could be reduced during analysis. For example, assume that a control on a plate, on which the prediction model is based (e.g. the positive product control), does fail the above proposed procedure. In the case of the positive product control, this would mean that none of the samples tested on that plate could be classified by the prediction model.

Furthermore, the robustness of the procedure with regard to systemic errors is noteworthy. It will work, even if the ELISA-plate is of low quality, e.g. with regard to coating, or if moderate systemic handling errors are present.

Application of the procedure to the available data

assay	CV_{max}	number of sets	Outlier procedure		Dixon test
			outliers	outlying sets	
THP-Bern	0.45	138	1	4	7
THP-Inns.	0.2	112	-	-	3
MM6	0.2	129	1	1	5
PBMC	0.45	154	1	1	6
WBT-Konst.	0.45	138	-	1	3

Table 1: Results of the outlier procedure

The results in table 1 are very promising, but should not be overestimated as mentioned above. For example, the problematic first run from Oslo with the THP-Bern assay with regard to Haemate can easily be handled with the proposed procedure, because three of the Haemates would have to be retested. In contrast, the Dixon test alone would not have detected any outlier in the three Haemates.

As can be seen in table 1, a harmonised choice for CV_{max} was sought. Alternatively, the more conservative $CV_{max} = 0.25$ for the two-plate cell line assays (THP-Innsbruck, MM6) could have been applied giving very similar results. The more restrictive $CV_{max} = 0.4$ for the THP-Bern and the two methods based on fresh blood could have also been chosen.

Unfortunately, the variation within sets of replicates for the WBT-NIBSC increased from Phases A and B to the pre-validation and is fortunately decreasing at the moment due to changes in the SOP. But considering the variation shown in Phases A and B and the harmonising aspect of the above proposal, a CV_{max}(WBT-NIBSC) = 0.45 seems to be appropriate.