

**Independent Peer Review Panel Report:
Five *In Vitro* Test Methods Proposed for Assessing Potential Pyrogenicity
of Pharmaceuticals and Other Products**

April 2007

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

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

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

**This document is available electronically at:
<http://iccvam.niehs.nih.gov/docs/pyrogen/PrRevPanFinRpt.pdf>**

The findings and conclusions of this report are those of the Independent Scientific Peer Review Panel and should not be construed to represent the official views of ICCVAM or its member agencies.

TABLE OF CONTENTS

***IN VITRO* PYROGENICITY INDEPENDENT PEER REVIEW PANEL iv**
PREFACE v
EXECUTIVE SUMMARY vii

A. REVIEW OF THE VALIDATION STATUS OF *IN VITRO* PYROGEN TEST METHODS 1

1.0 Introduction and Rationale for the Proposed Use of *In Vitro* Pyrogen Test Methods..... 19

1.1 Introduction 19

1.2 Regulatory Rationale and Applicability..... 20

1.3 Scientific Basis for the *In Vitro* Pyrogen Test Methods 22

2.0 *In Vitro* Pyrogen Test Method Protocol Components 24

2.1 Overview of How the *In Vitro* Pyrogen Test Methods are Conducted..... 24

2.2 Description and Rationale for the Test Method Components for Proposed Standardized Protocols 24

2.3 Basis for Selection of Test Method Systems 26

2.4 Proprietary Components 26

2.5 Number of Replicates..... 26

2.6 Modifications to the Test Method Protocols Based on ECVAM Validation Study Results 27

3.0 Substances Used for the Validation of *In Vitro* Pyrogen Test Methods..... 27

3.1 Rationale for the Substances or Products Selected for Testing..... 27

3.2 Number of Substances 27

3.3 Identification and Description of Substances Tested..... 28

3.4 Sample Coding Procedure 28

4.0 *In Vivo* Reference Data for the Assessment of Test Method Accuracy..... 28

4.1 Description of the Protocol Used to Generate *In Vivo* Data 28

4.2 Reference Data Used to Assess *In Vitro* Test Method Accuracy 29

4.3 Availability of Original Records for the *In Vivo* Reference Data..... 29

4.4 *In Vivo* Data Quality 29

4.5 Availability and Use of Toxicity Information from the Species of Interest..... 29

4.6 Information on the Relevance and Reliability of the *In Vivo* Test Methods 30

5.0 Test Method Data and Results 12

5.1 Test Method Protocol..... 30

5.2 Availability of Copies of Original Data Used to Evaluate Test Method Performance..... 31

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data 31

5.4 Summary of Results..... 32

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines..... 33

5.6 Lot-to-Lot Consistency of Test Substances..... 33

6.0 Relevance of the *In Vitro* Pyrogen Test Methods 34

6.1 *In Vitro* Pyrogen Test Method Relevance 34

6.2 Summary of the Performance Statistics for *In Vitro* Pyrogen Test Methods 34

7.0 Reliability of the *In Vitro* Pyrogen Test Methods..... 35

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of *In Vitro* Pyrogen Test Methods..... 35

7.2 Analysis of Repeatability and Reproducibility 35

7.3 Historical Positive and Negative Control Data 37

8.0 Test Method Data Quality..... 37

8.1 Adherence to National and International GLP Guidelines..... 37

8.2 Data Quality Audits..... 37

8.3 Impact of Deviations from GLP Guidelines 37

8.4 Availability of Laboratory Notebooks or Other Records..... 38

9.0 Other Scientific Reports and Reviews 38

9.1 Have Relevant Data Identified in Other Published or Unpublished Studies Conducted Using the *In Vitro* Test Methods Been Adequately Considered?..... 38

9.2 Are the Conclusions Published in Independent Peer-Reviewed Reports or Other Independent Scientific Reviews of the *In Vitro* Test Methods Adequately Discussed and Compared? 38

9.3 Are There Other Comparative *In Vitro* Test Method and RPT Data That Were Not Considered in the ICCVAM Draft BRD, But are Available for Consideration?..... 38

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement) 39

10.1 How the Five *In Vitro* Pyrogen Test Methods Will Refine, Reduce, or Replace Animal Use..... 39

10.2 Requirement for the Use of Animals 39

11.0 Practical Considerations 39

11.1 Transferability of the *In Vitro* Pyrogen Test Methods..... 39

11.2 Personnel Training Considerations 40

11.3 Cost Considerations 40

11.4 Time Considerations 40

12.0 Recommended Additional References 23

13.0 Summary of Validation Status of the *In Vitro* Pyrogen Test Methods 25

B.	REVIEW OF ICCVAM DRAFT TEST METHOD	
	RECOMMENDATIONS.....	26
1.0	Proposed Test Method Usefulness and Limitations.....	26
2.0	Proposed Test Method Standardized Protocols.....	26
3.0	Proposed Test Method Performance Standards.....	29
4.0	Proposed Additional Studies.....	30
C.	OVERALL PEER REVIEW OUTCOMES	33

IN VITRO PYROGENICITY INDEPENDENT PEER REVIEW PANEL

Karen Brown, Ph.D., (Panel Chair), DRL Pharma and Pair O'Doc's Enterprises, Parkville, Missouri

Brian Crowe, Ph.D., Baxter Vaccine AG, Orth an der Donau, Austria

Nancy Flournoy, Ph.D., University of Missouri-Columbia, Columbia, Missouri

Ihsan Gursel, Ph.D., Bilkent University, Bilkent, Ankara, Turkey

Ken Ishii, MD, Ph.D., ERATO, Japan Science and Technology Agency, Osaka University, Osaka, Japan

Jack Levin, MD, University of California-San Francisco, San Francisco, California

Albert Li, Ph.D., In Vitro ADMET Laboratories, Rockville, Maryland

David Lovell, Ph.D., University of Surrey, Guilford, United Kingdom

Melvyn Lynn, Ph.D., Eisai Medical Research, Inc., Ridgefield Park, New Jersey

Anthony Mire-Sluis, Ph.D., AMGEN, Inc., Thousand Oaks, California

Jon Richmond, MD, Home Office, Tayside, United Kingdom

Peter Theran, V.M.D., Massachusetts Society for the Prevention of Cruelty to Animals, Novato, California

Kevin Williams, Eli Lilly, Indianapolis, Indiana

PREFACE

This document is an independent report of the *In Vitro* Pyrogenicity Peer Review Panel ('Panel') evaluation of the validation status of five *in vitro* test methods for pyrogenicity testing. The Panel was convened as a National Institutes of Health (NIH) Special Emphasis Panel by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) to provide advice to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). This report summarizes the discussions, conclusions, and recommendations of the Panel's public meeting convened at the NIH in Bethesda, MD on February 6, 2007. ICCVAM and the ICCVAM Pyrogenicity Working Group (PWG) will consider the Panel report, along with comments from the public and the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM), and prepare final test method recommendations for U.S. Federal agencies. ICCVAM test method recommendations will be forwarded to U.S. Federal agencies for consideration and action, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. 285f-3, available at http://iccvam.niehs.nih.gov/docs/about_docs/PL106545.pdf).

The Panel considered five *in vitro* test methods submitted to ICCVAM by the European Centre for the Validation of Alternative Methods (ECVAM), a unit of the Institute for Health and Consumer Protection (IHCP) at the European Commission's Joint Research Centre. ECVAM submitted background review documents (BRDs) for these test methods to ICCVAM for consideration as replacements for the rabbit pyrogen test (RPT) in June 2005. The proposed test methods are:

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

ICCVAM established an ICCVAM PWG to work with NICEATM to carry out the evaluation of these test methods. The ICCVAM PWG developed draft test method recommendations and questions for consideration by the Panel. The ICCVAM PWG also collaborated closely with ECVAM throughout the evaluation process to obtain additional information for consideration by the Panel and ICCVAM.

The Panel was provided a comprehensive draft BRD prepared by NICEATM in conjunction with the PWG and ICCVAM. The draft BRD provided all available data and information related to the five *in vitro* pyrogen test methods. The five ECVAM submitted BRDs (one for each test method), the ECVAM response to PWG questions, and other supplemental information (i.e., key references and testing guidelines/regulations for pyrogenicity testing) were appended to the draft BRD. All of the information provided to the Panel was also made

publicly available, and public comments were requested via a *Federal Register (FR)* notice (Vol. 71, No. 238, pp. 74533-74534, 12/12/06). The *FR* notice also announced the public ICCVAM independent peer Panel review meeting scheduled for February 6, 2007.

The Panel was charged with:

- Reviewing the ICCVAM draft BRD for completeness and to identify any errors or omissions in the draft BRD
- Evaluating the information in the draft BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003¹) have been appropriately addressed
- Considering the ICCVAM draft test method recommendations for the following and commenting on the extent to which they are supported by the information in the draft BRD:
 - proposed test method uses
 - proposed recommended standardized protocols
 - proposed test method performance standards
 - proposed additional studies

At the Panel's public meeting on February 6, 2007, the Panel made recommendations for corrections and additions to the draft BRD and then discussed the current validation status of these five *in vitro* test methods. The Panel also commented on the ICCVAM draft test method recommendations for proposed test method uses, recommended standardized protocols, test method performance standards, and additional studies. The public was provided the opportunity to comment several times during the meeting. The Panel considered these comments as well as public comments submitted in advance of the meeting before concluding their deliberations.

The Panel gratefully acknowledges the efforts of NICEATM staff in coordinating the logistics of the peer review Panel meeting and in preparing materials for the review. The Panel also thanks Dr. Thomas Hartung (Head of ECVAM) for providing an overview of the test methods and for additional clarifications at the meeting. Finally, as Panel Chair, I want to thank each Panel member for their thoughtful and objective review of these test methods.

Karen Brown, Ph.D.
Chair, *In Vitro* Pyrogenicity Peer Review Panel
April 2007

¹ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC. NIEHS. The guidelines can be obtained at http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.htm.

EXECUTIVE SUMMARY

This report describes the conclusions and recommendations of the *In Vitro* Pyrogenicity Peer Panel ('Panel') regarding the validation status of five *in vitro* pyrogen test methods¹, and the ability of these test methods to individually serve as a substitute for the Rabbit Pyrogen Test (RPT) for the identification of Gram-negative endotoxin on a case-by-case basis, subject to product specific validation. The test methods are:

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

Panel Recommendations for the ICCVAM Background Review Document

The Panel stated that, in general, the information presented in the ICCVAM draft Background Review Document (BRD) was sufficient for its purpose. Exceptions are included within the body of the Panel report. The Panel identified a number of sections where clarification or a more comprehensive explanation would improve the ICCVAM draft BRD. For example, the extent to which the RPT is currently performed when risk assessments and regulatory decisions are concerned only with the presence of endotoxin should be provided. Likewise, a more detailed review of the various mechanisms and processes thought to be involved in the actual induction of fever itself, and a more detailed description of the statistical approaches used to evaluate the resulting data would be helpful. The Panel stated that the rationale for the selected test substances was neither appropriate nor acceptable and they recommended the inclusion of non-endotoxin pyrogens, protein- and lipid-containing materials that are used parenterally, and 'classical' examples of biological products and medical devices. The Panel also requested that the formal validation statement from the ECVAM Scientific Advisory Committee (ESAC) (and the supporting documents) be appended to the ICCVAM BRD. The Panel agreed that a comprehensive summary of findings on overall conclusions about the usefulness and limitations of each of the *in vitro* pyrogen tests compared to the Bacterial Endotoxin Test (BET) or the RPT should be included in the ICCVAM final BRD.

With regard to animal welfare, the Panel suggested that the ICCVAM final BRD provide information on the number of rabbits used for pyrogenicity testing to permit an accurate assessment of the actual impact on animal use. The Panel recommended that the ICCVAM

¹These test methods are referred to in this report as *in vitro* pyrogen tests in order to maintain consistency with the designation provided by the test methods' submitter (ECVAM). However, the Panel noted that this designation may be inappropriate because the usefulness and limitations for these test methods have been defined only for their ability to detect bacterial endotoxin and not other pyrogens.

final BRD discuss the practice of, and the U.S. Federal restrictions on, the reuse of rabbits in pyrogenicity testing, as well as the availability and use of the recombinant clotting factor C (rFC) that could replace the need for horseshoe crab hemolymph. The Panel also felt that the lack of direct parallel testing in rabbits with the products tested in the validation study was a significant limitation to the study design.

The Panel concluded that the cost and logistical considerations involved in conducting a study using the *in vitro* test methods were incompletely stated. The Panel recommended that a more detailed cost comparison for conducting the RPT and the *in vitro* test methods be performed. The Panel also commented that both the cost and logistical problems associated with the need to harvest and use human blood in four of the test methods were understated.

Validation Status of the *In Vitro* Pyrogen Test Methods

The Panel agreed that the applicable validation criteria have been adequately addressed in the ICCVAM draft BRD in order to determine the usefulness and limitations of these test methods to serve as a substitute for the RPT, for the identification of Gram-negative endotoxin on a case-by-case basis, subject to product specific validation. However, the Panel generally agreed that the performance of these test methods in terms of their reliability and relevance did not support this proposed use. A minority opinion (Dr. Peter Theran) suggested that the qualification in the above statement (i.e., that uses were subject to product specific validation) should allow for these test methods to be used for the specified purpose. A second minority opinion (Drs. Karen Brown, Albert Li, and Jon Richmond) expressed concern that it is not clear that the qualification included in the above statement would preclude the use of the *in vitro* test methods as replacements for the RPT in those circumstances where the BET is currently serving to replace the RPT.

Review of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Draft Recommendations for Test Method Usefulness and Limitations

The Panel concluded that the available data and demonstrated performance of these five *in vitro* test methods, in terms of their relevance² and reliability³, did not support the ICCVAM draft recommendations in terms of their usefulness and limitations. The Panel felt that the usefulness of these test methods for detecting Gram-negative endotoxin has not been properly assessed for concordance with the RPT or for relevance in comparison to the BET, and therefore, it was not possible to truly assess their usefulness and limitations.

One minority opinion stated (Dr. Peter Theran): This Panel has considered the failure to undertake additional RPTs a significant flaw in this validation study and therefore proposed that, in the future, similar validation studies should use the RPT to provide concordance data. I have no objection to the performance of *in vitro* tests in parallel with rabbit tests, which are already scheduled to be performed, in order to achieve concordance data. But, it is my

²The 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.

³A 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.

opinion, that any recommendation for *de-novo* parallel RPT should be accompanied by a statement, as follows: “The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis).” The inclusion of this statement would reinforce the importance of the 3R’s and would serve as a reminder of U.S. Federal law.

Review of the ICCVAM Draft Recommendations for Test Method Standardized Protocols

The Panel agreed that the information provided in the ICCVAM draft BRD supported the ICCVAM draft recommended protocols for these five *in vitro* test methods, providing that the list of inadequacies⁴ identified by the Panel with respect to reliability and relevance are fully addressed.

Review of the ICCVAM Draft Recommendations for Test Method Performance Standards

The Panel did not support the statement that the available data and demonstrated performance in terms of relevance and reliability supported the ICCVAM draft recommendations for these *in vitro* test methods in terms of their performance standards. The Panel noted several inadequacies with regard to the essential test method components for each *in vitro* test method and agreed that the demonstrated performance of certain aspects of several of the assays, particularly in terms of relevance, yielded some concern. With regard to the minimum list of reference substances, the Panel agreed that if the intent of the proposal was to replace the RPT with one or more of the *in vitro* test methods under consideration, then the *in vitro* test methods must be validated for all classes of substances (e.g., pharmaceuticals, biologicals, and implants) and medical devices that are tested with the RPT.

The same minority opinion directed towards the issue of parallel testing using the RPT as detailed above was expressed.

Review of the ICCVAM Draft Recommendations for Future Studies

The Panel agreed that to better determine the relevance of these *in vitro* test methods, the proposed additional studies should be performed using the ICCVAM proposed protocols, taking into account the Panel's comments and recommendations. The Panel also agreed that if the intended use of the *in vitro* assays were only to detect Gram-negative endotoxin, it would seem critical to include parallel studies with the BET in any future validation efforts. However, if the intended use of the *in vitro* methods is to evaluate substances containing endotoxin that are unable to be evaluated with the BET, then the parallel testing studies should include the RPT. The Panel recognized that these test methods could be applicable to a wider range of pyrogens and test materials, provided that they are adequately validated for such uses.

⁴Based on the list of 20 separate inadequacies outlined in this report, three Panel members felt that this list would be better described as a list of "many and substantial" inadequacies.

The same minority opinion directed towards the issue of parallel testing using the RPT as detailed above was expressed.

OVERALL PEER REVIEW OUTCOMES

This international independent Peer Review Panel, consisting of 13 expert scientists from five different countries, provided comments and recommendations on the usefulness and limitations of five *in vitro* pyrogen test methods for the detection and quantification of Gram-negative endotoxin and on the ICCVAM draft test method recommendations on the use of these *in vitro* methods as partial replacements for the RPT. These remarks are summarized below.

- In general, the information presented in the ICCVAM draft BRD was sufficient for the purpose of determining the usefulness and limitations of these test methods for their proposed use and for adequately addressing the applicable validation criteria on the basis of the currently available evidence.
- The available data and demonstrated performance in terms of their reliability and relevance do not at this time support the ICCVAM draft proposed use for these test methods (i.e., as a partial substitute or replacement for the RPT, for the identification of Gram-negative endotoxin, on a case-by-case basis, subject to product specific validation). To better characterize the test methods and more clearly define their reliability and relevance, the Panel recommended that specific additional studies be performed using the ICCVAM proposed protocols, taking into account the Panel's comments and recommendations.
 - The lack of parallel testing in the *in vitro* tests and the RPT, and the resulting lack of concordance data, was considered to be a major limitation of the validation study design. For this reason, the Panel recommended that future studies include parallel testing. A minority opinion (Dr. Peter Theran) associated with parallel testing was expressed as follows: "The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis)".
- The available data and demonstrated performance in terms of their reliability and relevance does not support the ICCVAM draft performance standards for these *in vitro* test methods for regulatory purposes.
- The information provided in the ICCVAM draft BRD supports the ICCVAM draft recommended protocols for these five *in vitro* test methods, providing that the list of inadequacies⁵ identified by the Panel with respect to reliability and relevance are fully addressed.
- These test methods could be applicable to a wider range of pyrogens and test materials, provided that they are adequately validated for such uses.
- It is critical to recognize, despite concerns about the performance of these five *in vitro* test methods, that a formal process exists for materials regulated under 21 CFR 610.9

⁵Based on the list of 20 separate inadequacies outlined in this report, three Panel members felt that this list would be better described as a list of "many and substantial" inadequacies.

to qualify these *in vitro* methods for the identification of Gram-negative endotoxin on a case-by-case basis, subject to product specific validation.

A. REVIEW OF THE VALIDATION STATUS OF *IN VITRO* PYROGEN TEST METHODS

1.0 INTRODUCTION AND RATIONALE FOR THE PROPOSED USE OF *IN VITRO* PYROGEN TEST METHODS¹

1.1 Introduction

1.1.1 Is the historical background provided for the *in vitro* pyrogen test methods and the rationale for their development adequate?

Yes, the Preface, the Executive Summary and **Section 1.1.1** of the ICCVAM draft Background Review Document (BRD) are all informative, clear, and concise with the following exceptions:

1. The action of pyrogens on circulating cells and the mechanism by which the pro-inflammatory cytokines produce pyrexia should be considered in **Section 1.1.1** instead of in **Section 1.3.2**).
2. The reduction in the use of animals to test medicinal products produced under current Good Manufacturing Practices (GMP) is an obvious goal. However, no information is provided on the current use of the Rabbit Pyrogen Test (RPT) or the bacterial endotoxin test (BET) (i.e., the approximate number of rabbits and horseshoe crabs used each year for pyrogen testing), or of anticipated trends in their use, or of the extent to which the RPT is currently used in contexts where risk assessments deem endotoxin to be the only relevant contaminant.
3. On lines 694-696 of the ICCVAM draft BRD (December 1, 2006), it is stated that the proposed *in vitro* tests were selected for their ability to replace the RPT. In the previous paragraph, it is stated that the RPT is capable of detecting both endotoxin and non-endotoxin pyrogens. Elsewhere, it is noted that these *in vitro* tests have not been validated for detecting non-endotoxin pyrogens. If the aim of testing these materials with the RPT is to detect a range of pyrogens, then these assays cannot, on the basis of information supplied in the validation dossier, completely replace the RPT.
4. A more detailed review of the various mechanisms and processes thought to be involved in the actual induction of fever itself, particularly in the case of drugs that are not administered intravenously, would have been useful. A number of reviews on this subject describe a far more complex picture than presented. These additional references include:
Netea et al. (2000) and Saper and Breder (1994).

¹These test methods are referred to in this report as *in vitro* pyrogen tests in order to maintain consistency with the designation provided by the test methods' submitter (ECVAM). However, the Panel noted that this designation may be inappropriate because the usefulness and limitations for these test methods have been defined only for their ability to detect bacterial endotoxin and not other pyrogens.

1.1.2 Is the previous review of the ECVAM validation studies adequately summarized?

Yes, the previous review of the ECVAM validation studies was adequately summarized. The questions resulting from the initial review have been answered and included in the ICCVAM draft BRD. However, it would have been better if the actual ESAC validation statement in full had been appended, as well as any documents used to support the ESAC conclusion. The ECVAM BRDs (though not the ESAC statement) contain inconsistent text relating to the possible practical uses of the novel tests that the validation tests were intended to support.

1.2 **Regulatory Rationale and Applicability**

1.2.1 Are the current regulatory testing requirements and ICCVAM prioritization criteria adequately discussed and up-to-date?

Yes, the current United States (U.S.) and European Union (EU) regulatory testing requirements are properly referenced and the relevant documents have been supplied. The previous product specific acceptance of peripheral blood mononuclear cell (PBMC) data by the U.S. Food and Drug Administration (FDA) is also mentioned in the Executive Summary.

Inclusion of the following information would have been useful:

1. It should be stated whether the acceptance of the PBMC data by the FDA was a replacement for the BET or the RPT. The document 21 CFR 610.9 provides for the use of alternative methods to test for pyrogenic substances as long as the use of these methods does not compromise the safety, purity or potency of the product. The 1987 FDA guideline on the validation of the BET as an end-product endotoxin test for human and animal parenteral drugs also sets forth acceptable conditions for the use of the test *in lieu* of the RPT. However, no mention is made of the fact that the European Directorate for the Quality of Medicines (EDQM) also has a working party of experts (apparently independent of ECVAM and ESAC) reviewing the whole area of *in vitro* pyrogens tests and their potential use.
2. The ICCVAM final BRD should discuss the availability and use of the rFC that could replace the need for horseshoe crab hemolymph.
3. The ICCVAM draft BRD gives few insights into how any recommendations, following acceptance by the relevant agencies, would be incorporated into U.S. Pharmacopeia (USP) and European Pharmacopeia (EP) test requirements.

Specific comments on the five ICCVAM prioritization criteria outlined in the ICCVAM draft BRD:

Criterion 1 (Applicability to regulatory testing needs and multiple agencies/programs): It is clear that the test methods are relevant to the end-product testing of a variety of healthcare products (for endotoxin) and that the FDA is the principal U.S. regulator for such products.

Criterion 2 (Warranted, based on extent of expected use or application and impact): It is clear from the documents that this criterion is only met with respect to the detection of endotoxin.

Criterion 3 (Potential to address any/all of the 3Rs): The tests have the potential to reduce or replace animal use and the associated morbidity and mortality. However, no information is provided in the ECVAM BRDs or in the ICCVAM draft BRD to permit the actual impact on animal use to be accurately assessed.

Criterion 4 (Potential to provide improved prediction): The documents indicate that the level of protection provided by each of the *in vitro* test methods is equivalent to that provided by the RPT. However, in the original ECVAM BRDs, it is recognized that sensitivity may have been underestimated and specificity overestimated as a consequence of having one of the spiked-sample points set at the regulatory limit. On lines 777-784 of the ICCVAM draft BRD (December 1, 2006), the statement that these methods would better predict the human pyrogenic response than the RPT because they use human cells is not supported by test results in the ICCVAM draft BRD. In contrast, it is stated on lines 1299-1303 of the ICCVAM draft BRD (December 1, 2006) that the pyrogenic response to endotoxin in rabbits and humans is “similar in both species. Based on these studies, the rabbit is considered to be predictive of the human response (and may often overpredict the response).”

Criterion 5 (Other advantages): The new test methods clearly take longer to produce definitive results. However, no animal facility is required. It was a surprise (in the absence of definitive cost information) that the novel tests were considered to be potentially more expensive than the RPT. Contract research organizations should be consulted on potential cost comparisons, as wide acceptance of these methods may in part be cost-dependent.

1.2.2 Is the description of the intended uses of the *in vitro* pyrogen tests complete?

These methods are proposed as partial replacements for the RPT. The RPT detects both endotoxin and non-endotoxin pyrogens, but the *in vitro* pyrogen tests have not been validated for non-endotoxin pyrogens. Therefore, they cannot be considered complete replacements for the RPT.

It is not clear when, or in which situations, the *in vitro* pyrogen test methods would be appropriate for use. The BET detects endotoxin in most cases and is used instead of the RPT for this purpose. The application of the *in vitro* test methods for the detection of endotoxin in sample types that cannot be measured in the BET is plausible; however, this proposed use would represent a very limited application for the *in vitro* pyrogen tests.

1.2.3 Are the similarities and endpoints measured by the proposed test methods and the reference (RPT) test method adequately described and discussed?

Yes, although the exact causes of the endpoint of the RPT (i.e., fever) are relatively complex and unclear, it has been known for many years that cytokines, especially those involved in the inflammatory response (i.e., IL-1, IL-6, and TNF) can induce febrile reactions. The development of tests based on the production of such cytokines from human white blood cells or cell lines appears to correlate well with the induction of fever in both the RPT and humans. However, the RPT detects a whole organ/body fever response; whereas, the proposed test methods detect only cytokine secretion. Evidence to suggest that detection of IL-1 or IL-6 is necessarily an indication of a febrile reaction is lacking. Additional information should be included in the ICCVAM final BRD on the relationship between IL-1

or IL-6 levels produced in cultures of monocytes and the development of fever in humans. The fact that the cytokine profiles for different endotoxins may vary between rabbits and humans should also be considered.

1.2.4 Is the description of the use of the proposed test methods in an overall strategy of hazard or safety assessment adequate?

Yes, the utility of the *in vitro* pyrogen methods as an addition to the current RPT, especially where non-endotoxin pyrogens are involved, has been clearly discussed. No specific claims are made for an immediate replacement of the RPT, although future studies may lead to such an event. The overall demonstration of the applicability of the methods to non-endotoxin detection is a stated goal. However, this goal does not appear to adequately match the methods employed since non-endotoxin standards were not used. One information gap (in the ICCVAM draft BRD and ECVAM BRDs) is the extent to which the RPT is currently performed when risk assessments and regulatory decisions are concerned only with the presence of endotoxin (that is clearly the intention when only the BET is used). Product-by-product validation will be required and the full extent of materials for which the new tests are not suited remains to be defined.

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

1.3.1 Is the purpose and mechanistic basis of the *in vitro* test method(s) adequately described and compared to known and/or suspected mechanisms/modes of action for fever production in humans?

Yes, the purpose and mechanistic basis of the *in vitro* test methods appears to be adequate while acknowledging that, at this point, the reference standard included in the validation study was Gram-negative endotoxin only. The mechanisms underlying fever induction, including the production of cytokines involved in the inflammatory cascade, appear to be important. The administration of such cytokines can directly induce fevers and their levels have been shown to dramatically increase during fevers. However, the known and suspected mechanisms/mode of action of fever may be far more complex than that described (see also response to **Section 1.1.1**).

The claim in **Section 1.3.1** to 'identify pyrogens' should perhaps be restated to 'detect pyrogens.'

1.3.2 Are the known similarities and differences of modes of action between the *in vitro* pyrogen test methods and the fever response in human and/or rabbits adequately considered?

Yes, an extensive literature search was performed that covered a wide range of cases illustrating the similarities and differences between the modes of action between the RPT, the *in vitro* pyrogen tests, and the induction of fever in humans. The correlation, or lack thereof, between the tests and human fever induction has been discussed in a scientifically valid manner. It should be noted that the RPT has served as a good predictor of human pyrogen response. Although there are false positives and false negatives associated with the RPT, it is not clear that these proposed *in vitro* assays provide better, similar, or worse results. A major concern is the lack of validation of these new assays directly compared to the RPT.

The mode of action is oversimplified. See response to **Section 1.1.1**, especially the reference to Netea et al. (2000) that provides an excellent review on the multiple-pathway mechanisms

that link cytokine responses (some of which are monitored by the proposed *in vitro* assays) and fever production. Furthermore, no description of the mode of action at either the molecular or cellular level is presented, which may prevent an adequate comparison between the methods. Specific questions that should be addressed include:

1. Are there any scientific data that compare IL-1 and IL-6 production and fever response between humans and rabbits?
2. Is the induction of IL-1 and IL-6 (or even fever) similar or different between endotoxin and other known TLR-4 ligands?
3. What is the mechanism of action for pyrogens that do not utilize TLR-4?
4. TLR-4-mediated IL-1 mRNA induction and the consequent release of mature IL-1 from cells by stimuli other than pyrogens are regulated by different molecular mechanisms. Are these mechanisms similar or different *in vitro* and *in vivo*, or between humans and rabbits?

1.3.3 Is the range of substances amenable to the *in vitro* pyrogen test methods, and are the limits of the test methods adequately characterized?

Yes, given what is known of materials with the potential to interfere with the test system supplemented by the need for product-by-product validation and the exclusion of interference. More work will have to be carried out to understand the types of materials that could be tested in these assays and how they would be handled (e.g., cell therapies and implants). However, it must be considered that a manufacturer of a medicinal product would have to validate the *in vitro* method they have selected specifically for their particular product before it would be acceptable to any regulatory authority. Thus, comprehensive testing of a wide variety of substances may not be necessary to introduce these tests into general use. Insufficient information exists at present to be confident that all types of materials that will demonstrate interference have been identified (e.g., materials that are cytotoxic, contain immunological adjuvants, or have antipyretic properties) but case-by-case evaluation provides the necessary safeguards. In addition, although the test methods have been shown to have the potential to identify non-Gram negative pyrogens, the validation study only presented detailed data and analysis with respect to the tests' potential to detect Gram-negative endotoxins (see also **Section 1.3.1**).

However, with respect to the limits of the test methods, no mention is made of the wide range of drugs that are toxic to blood cells or that induce a substantial pro-inflammatory response and consequently are not amenable to testing by these methods. Many pure, well-established non-endotoxin compounds have been shown to activate blood cells, including monocytic cells, to produce pro-inflammatory cytokines *in vitro* and *in vivo* (see suggested additional references [Ishii et al., 2005; Ishii and Akira, 2006] in **Section 12.0**).

On page 1-5, line 770 of the ICCVAM draft BRD (December 1, 2006) states, "Although the *in vitro* BET is performed using hemolymph (the equivalent of blood) drawn from *Limulus polyphemus* (horseshoe crabs), which are subsequently returned to the wild, there is some mortality associated with the procedure (which requires approximately 20% of the animal's total blood volume)". This concern has been largely solved with the commercial introduction of rFC, which was originally cloned from the horseshoe crab. This commercial product is currently being compared to the BET for submission for inclusion in the USP. A need for a

replacement for the RPT for early compound development testing and testing of biologics that have some propensity to harbor non-endotoxin pyrogens remains to be fulfilled. Thus, the goals of the overall effort need further refinement. Endotoxin is, of course, the important standard for validation purposes but non-endotoxin standards need to be characterized to further such a test for non-endotoxin testing; this concept is referred to on page 1-7, lines 821-822 of the ICCVAM draft BRD (December 1, 2006).

2.0 IN VITRO PYROGEN TEST METHOD PROTOCOL COMPONENTS

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

Are there gaps or missing information in the overview of how the tests are conducted?

This section seems adequate and complete. The overview of how the *in vitro* pyrogen tests are carried out is brief and to the point. The assays essentially expose human blood cells (either primary or cell line derived) to a test substance that may or may not induce cytokine release. Any cytokine release is subsequently detected with an immunoassay.

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

Are the description and rationale for each of the following test method protocol components for the recommended versions of the *in vitro* test methods adequately described and appropriate? Should any protocol components be modified, and, if so, why? Are additional protocol components needed, and, if so, why?

2.2.1 Materials, equipment and supplies

Specific concerns with respect to human blood donors include: diurnal variation, genetic polymorphisms (i.e., in genes coding for Toll-like receptors [TLRs], cytokine receptors, response elements, etc.), and number of donors required.

The effect of components in the blood and their effects on the assay systems are not clear (i.e., the effect of variations in the number of monocytes in peripheral blood, which range from 2 to 10%, as well as the effect of neutrophil or lymphocyte presence on the cytokine response).

2.2.2 Endpoint(s) measured

The viability of the human blood cells should be monitored before and after incubation with the test samples. Cytotoxic substances should not be tested with these methods.

2.2.3 Duration of exposure

A fixed exposure time rather than a broad range of exposure times (e.g., 16 to 24 hours) should be defined.

2.2.4 Known limits of use

It is suggested that the *in vitro* pyrogen tests are suitable for the testing of medical devices and materials by direct contact rather than testing extracts. However, direct contact may not adequately permit the solubilization or leaching of potential pyrogens.

2.2.5 Nature of the response assessed

The nature of the response assessed is accurately summarized. However, a description of the blood cell types known to respond to pyrogens by producing IL-1 and/or IL-6 should be included.

2.2.6 Appropriate negative, vehicle, and positive controls and the basis for their selection

The ECVAM BRDs do not discuss why high quality Gram-positive material (Lipoteichoic acid [LTA]) available from the University of Konstanz was not also used as a 'model' pyrogen. The inclusion of such non-endotoxin positive controls would be useful in future validation studies to further characterize the usefulness and limitations of these methods for the detection of such substances.

2.2.7 Acceptable ranges of negative, vehicle, and positive control responses and the basis for the acceptable ranges, or procedures for establishing acceptable ranges

The ECVAM BRDs indicate that (refer to **Sections 6.1.1**), with hindsight, the use of an endotoxin spike solution at the threshold pyrogen dose (marking the pass/fail level for regulatory purposes) was not wise. See above (response to Criterion 4, **Section 1.2**) regarding possible relevance to determination of sensitivity and specificity of the novel test methods.

2.2.8 Nature of the data to be collected and the methods used for data collection

The description of the nature of the data to be collected and the methods used for data collection is accurate.

2.2.9 Type of media in which data are stored

The type of data storage media seems to fit the purpose. However, one printed version of the data should be stored.

2.2.10 Measures of variability

The description of the measures of variability reflects the current state of knowledge. Other relevant physiological variables may exist but the main sources of potential variation seem to have been addressed.

2.2.11 Statistical or nonstatistical methods used to analyze the resulting data

Generally adequate, but additional clarification is desired. It would seem appropriate to use a consistent approach across assays. For example, in some places, Dixon's test was used to identify outliers, while in others Grubb's test was used; the reasons and contexts for these differences are not apparent. However, it is accepted that minor problems arise with the calculation of sensitivity and specificity of the novel test methods from using a spike-point coincident with the regulatory limit.

The statement that "using an endotoxin curve, the endotoxin content of the product is calculated" is not true. The *in vitro* pyrogen test is not specific for Gram-negative endotoxin

and therefore, it is impossible to know whether the response measured is due to endotoxin or another pro-inflammatory response reactive substance in the sample.

2.2.12 Decision criteria and the basis for the prediction model used to classify a test substance as positive or negative for the presence of a pyrogenic material

The RPT data used to set the pass/fail criteria were produced in one rabbit strain in one laboratory and were not obtained concurrently within the validation study.

It is not clear that the criteria used to assign test results as positive or negative are based on the precise criteria set out in the USP. The significance of any deviations from these criteria is also not clear.

2.2.13 Information and data that will be included in the study report and availability of standard forms for data collection and submission

The descriptions provide a good overview of each test for the purposes of comparing and contrasting them with one another and with current methods.

2.3 Basis for Selection of Test Method Systems

Is the description of the basis for selection of the test method systems complete and appropriate?

A brief description of the advantages of each test method have been provided and are appropriate for considering the limitations of the existing tests for pyrogens, namely the RPT and the BET.

2.4 Proprietary Components

Are proprietary components appropriately identified (if applicable), and are the procedures adequate for ensuring their integrity from 'lot-to-lot' and over time?

The licensing procedure and availability of the Mono Mac 6 (MM6) cell line is unclear. Variations in the MM6 cell line (and primary cells) must be properly controlled. A direct comparison of the commercially available enzyme-linked immunosorbent assay (ELISA) kits should also be included in the ICCVAM final BRD.

2.5 Number of Replicates

Are the numbers of replicate and/or repeat experiments appropriate for each test method?

The appropriate number of donors from which to collect blood cells is unclear. Furthermore, some of the test methods permit pooling of blood donors while others do not. The rationale for these differences is unclear.

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

Are the protocol modifications based on ECVAM validation study results appropriate for each modified test method?

Yes, only minor modifications were made to two of the five assays, both to improve assay performance, and therefore the limited explanations are appropriate.

3.0 SUBSTANCES USED FOR THE VALIDATION OF *IN VITRO* PYROGEN TEST METHODS

3.1 Rationale for the Substances or Products Selected for Testing

Is the rationale for the selected test substances appropriate and acceptable?

No, the only rationale given for the choice of test substances is that they represent marketed parenteral pharmaceuticals that were readily available at reasonable cost. According to their USP monographs, seven of the ten test substances are currently tested in the BET, not in the RPT. No USP monographs exist for the remaining three because pyrogen testing is not required. The inclusion of test substances that may interfere with the *in vitro* responses should be tested.

Although the test materials spiked with endotoxin are described as having been initially pyrogen-free and having been approved for clinical use, all that can be said with confidence is that they did not contain a level of pyrogen above the permissible or tolerable limit. As a result, in describing the concentration of endotoxin in the spiked sample, it is more correct to state the minimum level of endotoxin they were known to contain rather than offering an absolute value.

Non-endotoxin pyrogens should be evaluated because these pyrogens must be tested in the RPT and they cannot be tested in the BET. The list of test substances should also include protein- and lipid-containing materials that are used parenterally. No 'classical' examples of biological products or medical devices were included; thus, the validation for either of these categories has not been provided.

Although it is stated that endotoxin was chosen as a model pyrogen, insufficient information exists in the ICCVAM draft BRD or in the supporting ECVAM BRDs to support this claim. The validation study documents, the ESAC validation statement and the ICCVAM draft BRD claim only that the test methods are suited for the detection/qualification of Gram-negative endotoxin for regulatory testing.

3.2 Number of Substances

Please comment on the adequacy of the number of substances used in the performance analyses.

The total number of substances included in the validation study is adequate only for validation of a specific class of products. Replacement of the RPT would require a much larger number of substances because of the wide range of product classes that would require testing. Moreover, the test substances should have represented each of the major classes of

products normally tested in the RPT (e.g., medical devices, biologicals, implants, and those substances known to interfere with the RPT, the BET, and/or the *in vitro* pyrogen tests) as positive controls for interference testing.

3.3 Identification and Description of Substances Tested

Are the test substances adequately identified and described?

The samples included in the validation process are adequately identified and described such that they could be readily obtained for future studies. However, more information on their purity and batch/lot numbers is needed in order to adequately demonstrate that the same substances were tested throughout the validation studies. In response to a request for additional information, ECVAM did provide the lot numbers used in the validation study, which demonstrated that they were identical. However, some differences in the lots tested in the catch-up validation study were noted (e.g., two of the ten substances had different lot numbers due to the lack of availability; one was a different substance with the same active ingredient).

3.4 Sample Coding Procedure

Were the coding procedures used in the validation studies appropriate?

The coding procedures were adequate for the assessment of relevance during the validation studies. However, the identity of the substances used in the reproducibility analyses was not blinded (although the spike concentrations were). A reason was not given.

4.0 *IN VIVO* REFERENCE DATA FOR THE ASSESSMENT OF TEST METHOD ACCURACY

Are the *in vivo* reference data used in the validation study appropriate to allow for adequate assessment of test method relevance² (accuracy/concordance, sensitivity, specificity, false positive and false negative rates) of these *in vitro* pyrogen test methods as a partial replacement for the RPT, for materials which may be contaminated with gram-negative endotoxin, but which cannot be tested by the BET?

No, a summary of the reference data demonstrating whether substances that were shown to be pyrogenic in humans either passed or failed the RPT, BET or *in vitro* pyrogen tests would have been useful.

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

Is the RPT protocol used to generate reference data for the cited studies appropriate?

²The 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.

The RPT protocol and the pass-fail criteria used would not meet the current USP requirements. The significance of these deviations is not clear. The data are derived from a single study carried out at the Paul Ehrlich Institute (PEI) where historical controls tested over five years were accumulated and analyzed. The protocols used at the PEI were based on the EP monograph for the RPT, although this fact is not explicitly stated in the publications. Furthermore, the detailed protocol used by this laboratory was not provided.

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

Is the interpretation of the reference data used to assess *in vitro* test method accuracy correct? Is any other data or information needed to determine the accuracy of the test methods?

The reference data were previously and separately generated by one protocol, in one laboratory, using one strain of rabbit, and two sources of endotoxin. A second study, undertaken in Brazil, is cited. The response criteria of the Brazilian study do not match those of the PEI study. It is not clear why the Brazilian study was not relied upon for the validation study.

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

Are there any concerns with the availability of the original reference data records as described?

The data were derived from a single study at the PEI and presented in graphical form. No additional data were available for analysis. Archived records have not been audited by ECVAM or ICCVAM.

4.4 *In Vivo* Data Quality

Are there any concerns with the RPT data quality?

The ECVAM documentation is not sufficiently specific and in the absence of the primary data, the quality of the RPT data is unknown. The ICCVAM draft BRD does not clearly indicate the GLP status of the laboratory or of the study. However, the PEI did not have formal GLP accreditation (refer to **Section 5.5**, ECVAM response to a request for additional information).

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

Is the discussion of the availability of relevant pyrogenicity information for humans adequate and appropriate? Are there other sources of quality human data for pyrogenicity that should be considered? Would human data be compatible with regulatory needs (e.g., exposure duration, individual sensitivity)?

The available data are limited. However, the availability of information on clinical adverse events resulting from the administration of medical products producing pyrogenic effects, and the relevant pre-clinical test data, would be an excellent source of human data. See

suggested additional reference (McKinney et al. 2006). The data would reflect responses seen using appropriate human exposure; thus, it should be compatible with regulatory needs.

A discussion of relevant pyrogenicity information for humans is present in the ICCVAM draft BRD, but additional information is needed. An extensive literature on acute human pyrogenicity responses exists and this data should be better reviewed. Effects of longer exposure and individual sensitivity are available in Rylander (2002).

The data in the cited paper by Greisman and Hornick (1969)³ are not accurately described (page 4-6, lines 1299-1301 of the ICCVAM draft BRD [December 1, 2006]).

4.6 Information on the Relevance and Reliability of the *In Vivo* Test Methods

Is what is known about the relevance and reliability⁴ of the RPT adequately discussed and appropriately considered?

The appropriateness of the theoretical assumption model is unclear. It is not clear how the sensitivity and specificity values have been derived using this model. Therefore, reference to these values as accurate figures (particularly with respect to the 58% sensitivity) is a concern.

The theoretical sensitivity and specificity for the RPT that has been supplied does not seem to reflect its performance in practice or the regulatory decisions and level of patient safety that RPT data currently supports.

The 'correct' figures for the theoretical specificity of the RPT are confusing. It was stated to be 83% in **Section 4.6** of the ICCVAM draft BRD (December 1, 2006) but given as 88.3% in the ECVAM response to ICCVAM questions (page 24).

However, this difference has little bearing on the overall interpretation of the results.

5.0 Test Method Data and Results

5.1 Test Method Protocol

Are the *in vitro* test method protocols used to generate each set of data considered in the ICCVAM draft BRD appropriately described?

The following problems with all five *in vitro* test method protocols were noted:

1. Quality control (QC) testing of cell viability is not performed. Viability testing of the human cells before and after incubation should be performed.
2. No microscopic examination for anticipated levels of cell fragments and debris is described.
3. Substances should not be tested at cytotoxic concentrations by these methods.

³ Greisman SE, Hornick RB. 1969. Comparative pyrogenic reactivity of rabbit and man to bacterial endotoxin. Proc Soc Exp Bio Med 131(4):1154-1158.

⁴A 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.

4. More detailed source information and the pyrogen status (i.e., pyrogen-free) should be required for all protocol components.
5. A description of the procedure used for donor recruitment and donor selection is not provided.
6. A description of the protocols used for preparation of blood samples for the cytokine assays is not found.

The following problems with specific test method protocols were noted:

1. In the Cryo WB IL-1 assay, the incubation of the test sample is performed in the presence of 10% DMSO (methods for its removal after thawing of the cryopreserved cell preparation and before its use are not described). DMSO is known to effect the detection of certain cytokines. In response to a request for additional information, ECVAM indicated that the DMSO is not removed.
2. A limit to the passage number should be defined for the MM6 cell cultures.
3. The use of the terms RPMI-M and RPMI-C (described in the ECVAM MM6/IL-6 Standard Operating Procedure [SOP]) is confusing in the ICCVAM protocol.
4. A typographical mistake appears in the ICCVAM MM6/IL-6 protocol (lines 285 and 286 of the ICCVAM draft BRD [December 1, 2006]) where 'FBS' is stated instead of 'PBS'.

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

Has the availability of the original data use in the test method performance evaluation been adequately described?

Yes, the availability of the original source data has been adequately described.

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

Are the statistical and non-statistical approaches used in each cited study to evaluate the *in vitro* test results appropriate? What other approaches could have been used?

The statistical approaches appear adequate. However, it is suggested that more emphasis should have been placed on a quantitative estimate of pyrogen concentrations rather than dichotomizing results based upon a hypothesis test. One would have expected to see *a priori* criteria for successful validation in terms of acceptable performance statistics.

The term 'correlation' appears to be used colloquially (e.g., lines 1365 and 1373 of the ICCVAM draft BRD [December 1, 2006]); a correlation is not a percentage. Therefore, 'correlation' should be replaced with 'association' everywhere, except when Pearson's correlation is being referenced.

Information on the identification and elimination of aberrant data from **Section 4.2** of the Trial Data report should be included in the ICCVAM final BRD.

5.4 Summary of Results

Is the summary of the results for each test method appropriate and adequate?

No data were presented to confirm that results in the *in vitro* tests reflect human physiological responses or that production of IL-1 or IL-6 *in vitro* correlates with pyrogenicity *in vivo*. A quantitative link between IL-1 and IL-6 concentrations and their donor-to-donor variation with physiological effects was not presented. It should be mentioned that according to Schindler et al. (2006)⁵, which describes the validation of the Cryo WB/IL-1 method, testing problems existed with many of the products included in the study (up to 9 of 10). This is evident by failure of the positive product control (PPC), which under normal circumstances would invalidate the test. Instead, when the PPC failed, the authors report that the saline control was used in place of the PPC and the experiment was still considered acceptable. This practice is unacceptable.

The lack of direct parallel testing in rabbits with the products tested in the validation study prevents an evaluation of actual physiological effects. It also would have been of assistance to the Panel if information had been provided to document that the use of human cells could partially replace the BET and RPT for the detection of substances that are pyrogenic in humans.

Some of the data (or lack thereof) indicate significant limitations of the *in vitro* assays. Specific examples are listed below:

1. In the ICCVAM draft BRD (December 1, 2006), page 2-7, line 989: The use of a single donor in the WB/IL-1 assay is inadequate.
2. In the ICCVAM draft BRD (December 1, 2006), page 2-10, line 1050: There are no data offered to document that the use of human cells will better reflect human physiological responses or that production of IL-1 or IL-6 *in vitro* correlates with pyrogenicity *in vivo*.
3. In the ICCVAM draft BRD (December 1, 2006), page 6-2, line 1456: 20 of 150 runs in the Cryo WB/IL-1 assay were not usable. Even then, the false positive rate of the remaining 120 assays was 18.6%.
4. In the ICCVAM draft BRD (December 1, 2006), page 6-4, line 1493: 1 of the 3 validation laboratories had a 50% false positive rate for the PBMC/IL-6 assay.
5. In the ICCVAM draft BRD (December 1, 2006), page 7-7, Table 7-4: Agreement across three validation laboratories was only 57% for the WB/IL-1 assay.

⁵ Schindler S, Spreitzer I, Löschner B, Hoffmann S, Hennes K, Halder M, Brügger P, Frey E, Hartung T, Montag T. 2006. International validation of pyrogen tests based on cryopreserved human primary blood cells. *J Immunol Methods* 316:42-51.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

For each set of data for each test method, is whether coded substances were tested and whether experiments followed GLP Guidelines adequately documented?

The use of coded substances is adequately documented, but the rationale for not blinding the identity of the three substances tested in the reliability analyses is not known. The *in vitro* pyrogen test studies were conducted 'in the spirit of' GLP requirements. However, gaps and lapses in the information supplied by ECVAM would indicate that none of the testing laboratories were audited in real-time. In response to a request for additional information from ECVAM, it was stated that:

"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 readings into Excel sheets provided by the statistician. 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."

5.6 Lot-to-Lot Consistency of Test Substances

Is the information on the 'lot-to-lot' consistency of the test substances, the time frame of the various studies, and the laboratory in which the study or studies were conducted, adequately described?

Information on specific lots used in the validation studies was not provided in the ICCVAM draft BRD and therefore, lot-to-lot consistencies cannot be evaluated. Additional information has been received to demonstrate that the same lots were tested in the validation study, but there were lot differences in 2 of 10 substances used in the catch-up validation study. In addition, because one of the substances used in the original validation was no longer available, a different substance (with the same active ingredient) was used in the catch-up validation.

Unfortunately, little or no high concentration protein samples (e.g., Factor VIII concentrates or 5-25% human albumin samples), where lot-to-lot inconsistencies might be expected, were tested in the validation studies. This exclusion was explained to some extent by ECVAM in the responses that they provided to the ICCVAM/PWG questions. Interference testing for all sample types should be tested on multiple lots (also see the specific inadequacy [No. 10, lines 1361-1362] noted in the proposed test method standardized protocols).

6.0 RELEVANCE OF THE *IN VITRO* PYROGEN TEST METHODS

6.1 *In Vitro* Pyrogen Test Method Relevance

Has the relevance (e.g., accuracy/concordance, sensitivity, specificity, positive, and negative predictivity, false positive and false negative rates) of the *in vitro* test methods for detection of pyrogens, as defined by statutes in the United States Code (see Section 1), or for sterility testing defined by the U.S. Pharmacopeia or the International Standards Organization, been adequately evaluated? Are the discussions of the relevance of each *in vitro* test method and the reference test method appropriate and accurate?

In general, the evaluation of the relevance of the *in vitro* pyrogen tests appears to have been appropriately demonstrated and discussed, but limited by the ability to judge a positive versus negative response using a cut-off at 0.5 endotoxin units (EU)/mL. Furthermore, because only endotoxin-spiked samples were tested, relevance has been demonstrated only for the detection of bacterial endotoxin.

This section is entirely focused on comparisons between the *in vitro* pyrogen test methods since the RPT was not carried out in parallel, but rather estimates of the RPT performance were modeled statistically. The validity of this approach remains in question due to the nature of the RPT, where a definitive cut-off point does not exist, but was defined based on the results generated from the historical database. Therefore, no data exist with which to establish concordance with the RPT and thus, the discussion on concordance with the RPT is speculative.

Discrepancies between Table 6-1 and the accompanying text of the ICCVAM draft BRD (December 1, 2006) for the cryopreserved PBMC assay prevented assessment of this method.

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

Is the summary of the performance of the test methods adequately described? Are the strengths and limitations of each *in vitro* test method adequately identified?

A more critical description and explanation are needed (i.e., a failure of the prediction model or a failure of the assay to correctly detect the pyrogen concentration) for the cases where the test failed to correctly classify the pyrogen concentration.

The discussion of the strengths and limitations of each of the test methods should be expanded. Specific points include:

1. Inadequate performance is noted for: a) Cryo WB/IL-1 (false positive rate = 18.1%); b) WB/IL-1 (false negative rate = 27.3%); c) WB/IL-1 (false positive rate = 16.4%). High false positive rates are clearly a concern for manufacturers since lots may be unnecessarily withheld from release.
2. The high exclusion rate for individual runs in the case of the Cryo WB/IL-1 test (20% - 30% out of 150 runs) due to excessive variability among the four replicates, even with a relatively high coefficient of variation (CV) criteria (CV > 45%).

3. The low sensitivity (only 72%) for the WB/IL-1 assay, resulting in an extremely high false negative rate (27.3 %). High false negative rates would obviously be a major concern, as endotoxin-contaminated lots would be released.

Taken together, these statements could indicate that the WB/IL-1 assays (WB/IL-1 Cryo WB/IL-1, and WB/IL-1 96-well plate method) do not, in general, perform as well as the other assays that measure an IL-6 response.

It would have been very interesting to have had the opportunity to compare performance analysis data for the BET, since only endotoxin spiked samples were used in the validation and endotoxin testing is now the intended use for the *in vitro* pyrogen tests. Unfortunately, the BET was not performed in the validation so no direct comparison can be made between it and the new *in vitro* assays.

7.0 RELIABILITY OF THE *IN VITRO* PYROGEN TEST METHODS

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

Is the selection rationale and the number and types of substances used to evaluate the reliability of the *in vitro* test methods (intralaboratory repeatability and intra- and inter-laboratory reproducibility) as well as the extent to which the chosen set of substances represent the range of possible test outcomes appropriate?

The use of a standard material such as the endotoxin control (WHO-LPS, 94/580) is a valid choice for conducting the experiments described since it is a well-characterized, well-documented material. However, the rationale for the selection of the drugs used in the studies for evaluating reproducibility versus sensitivity/specificity is not clear, except that they were manufactured under GMP, were licensed products, were reported not to be contaminated with unacceptable levels of endotoxin, and were all available at reasonable cost. It would have been more appropriate to evaluate reliability using a subset of the drugs used in the sensitivity/specificity studies.

7.2 Analysis of Repeatability and Reproducibility

Are the analyses and conclusions regarding the intralaboratory repeatability and reproducibility and the intra- and inter-laboratory reproducibility of each test method appropriate? Should other analyses be considered?

The experiments performed to evaluate intralaboratory repeatability and intra- and inter-laboratory reproducibility were overly complicated. However, the analysis based on 'positive or negative' calls suggests that the reliability of these *in vitro* test methods are generally acceptable both within and between laboratories, although a more critical description is needed to explain the lack of agreement among some test results.

It is interesting that the variability of the cell line-based MM6 assay is much reduced compared to that obtained for the whole blood assays, although this observation did not translate into an improved ability to assign a negative or positive status to a sample.

The following deficiencies were noted:

1. More discussion is needed about the use of a coefficient of variation (CV) analysis to evaluate the reliability of the *in vitro* test methods, including how an 'acceptable' CV was identified (e.g., 45% in the WB/IL-1 assay) and why the criteria for an acceptable CV was inconsistent among the different *in vitro* test methods.
2. It is not clear which statistical test(s) was used to detect outliers and whether the test(s) was based on original or log-transformed data. Furthermore, it is not clear how many data points were identified as outliers and how they were subsequently handled during data analysis. The information provided by ECVAM addressing these concerns should be integrated into the ICCVAM final BRD.
3. A quantitative assessment of the intra- and inter-laboratory variability would have been more informative than an assessment based on dichotomizing the test results. The assessment should have included estimates of the amount of inter- and intra-laboratory variability and the number of replicates needed to estimate the sources of variability. Consistent with general practice, acceptable levels of variability should have been identified *a priori*, and it should have been recognized that formal hypothesis testing is essential with the alternative hypothesis being no difference between groups.
4. Potential problems related to plate-to-plate variation and/or other plate design issues should be addressed in the ICCVAM final BRD.
5. The use of the term 'mean value calculated' needs to be clarified.
6. It is misleading to state that the test substances were spiked at four concentrations when two of the spikes are at the same concentration. The concentrations should be noted explicitly, even in summaries if this is their first reference.
7. The ICCVAM final BRD should state whether or not the data were log-transformed prior to analysis (as was stated in the ECVAM BRDs). Furthermore, in the ECVAM BRDs, the decision rationale for performing a log transformation versus a square-root transformation of the data should be provided. In all ECVAM BRDs, it is not clear whether all analyses used log-transformed data or if transformed data were used only for the *t*-test in the classification phase of the analysis (e.g., ECVAM BRD for WB/IL-1, page 25).
8. The ECVAM BRDs state that all data are log-transformed, but the y-axis on the graphs is labeled OD 450 (e.g., ECVAM BRD for Cryo WB/IL-1, Appendix D). The data should be log-transformed if this has not yet been done. The CV after transformation is of most interest; however, the figures appear to give data before the transformation indicating that the variance increases with the mean. Data after the transformation should also be plotted to show that the relationship of the mean and the variance is well suited to the log transformation. The analysis with respect to the transformation needs to be

clarified. The values on the x-axis are unreadable and need to be given in the legends or in the description that accompanies each figure.

9. The notation used in the *t*-test (e.g., the subscripts on the population and sample means) needs to be defined. In the standard two-sample *t*-test, the groups are assumed to be independent. However, it looks like one group is a collection of subgroups and the other group is one of these (i.e., the data from one group are used in the calculation of both means). This point needs to be clarified.

7.3 Historical Positive and Negative Control Data

Is the availability of historical negative and positive control data adequately considered?

The fact that the *in vitro* pyrogen test methods are not in routine use except for the two manufacturers cited (who are unlikely to provide what would be considered proprietary data) leads to a paucity of historical data.

8.0 TEST METHOD DATA QUALITY

8.1 Adherence to National and International GLP Guidelines

Is the extent of adherence to national and international GLP guidelines for all submitted *in vitro* and *in vivo* test data and the use of coded substances and coded testing adequately presented?

It is clear that SOPs exist and that protocols were developed for all *in vitro* experiments performed. However, the precise GLP status of the studies and the test laboratories is not clearly stated and the ICCVAM final BRD should be revised to clarify this information. The *in vivo* data are derived from a single published study.

8.2 Data Quality Audits

Are the results of any data quality audits, if conducted, adequately summarized?

From the information provided, it would seem that no audits were undertaken while the studies were in progress. However, the ECVAM BRDs state that 'deviations' were recorded but no further details or information is provided. A summary of the GLP deviations that occurred would have been useful for determining their overall significance to the experimental outcome.

8.3 Impact of Deviations from GLP Guidelines

Does the lack of an evaluation of the impact of deviations from GLP guidelines affect the data analysis?

This question cannot be answered, as no data have been provided on any deviations from GLP guidelines.

8.4 Availability of Laboratory Notebooks or Other Records

Is the availability of laboratory notebooks or other records for an independent audit adequately discussed?

Yes, the study authors state that all raw data are available for inspection and have been archived appropriately.

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Have Relevant Data Identified in Other Published or Unpublished Studies Conducted Using the *In Vitro* Test Methods Been Adequately Considered?

Although an extensive literature has been cited and discussed, no attempt at a comprehensive summary of findings or overall conclusions about the relevance of the *in vitro* pyrogen tests compared to the BET or the RPT, or the advantages/capabilities or disadvantages/limitations of the individual *in vitro* assays, has been presented in the ICCVAM draft BRD.

The following additional references should be included (see **Section 12.0** for full citations):

1. Marth and Kleinhappl (2002). The studies described here indicate the importance of monitoring multiple pro-inflammatory cytokine responses. In the specific case cited, the TNF- α pro-inflammatory cytokine response appeared to correlate best with fever.
2. Norata et al. (2005), van Deventer et al. (2000), von Aulock et al. (2003) are relatively new studies that evaluate the effects of genetic polymorphisms on TLR-4 responses.
3. Martis et al. (2005). This paper describes a situation where the PBMC/IL-6 assay was used to help resolve a non-febrile adverse drug reaction issue with a licensed product.

9.2 Are the Conclusions Published in Independent Peer-Reviewed Reports or Other Independent Scientific Reviews of the *In Vitro* Test Methods Adequately Discussed and Compared?

Yes, the conclusions are adequate for the published data.

The formal ESAC validation statement and other EU validation expert/panel process documents should be appended to the ICCVAM final BRD.

9.3 Are There Other Comparative *In Vitro* Test Method and RPT Data That Were Not Considered in the ICCVAM draft BRD, But are Available for Consideration?

It is known that manufacturers have parallel test result data for the BET and RPT for specific products, which unfortunately are not published or peer reviewed. As a consequence, a number of companies are now advocating that they should be permitted to use the BET as an alternative to the RPT to detect the presence of Gram-negative endotoxin on a case-by-case basis, such as for testing of established products with documented proof that safe, reliable and consistent GMP production and QC procedures are in place.

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

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

Is the extent to which the *in vitro* test methods will refine (reduce or eliminate pain or distress), reduce, or replace animal use in the RPT adequately described?

No numbers are included regarding the current number of rabbits used and/or killed with this test. These estimates would be helpful when assessing the potential impact of these *in vitro* tests. However, given that the proposed use for these test methods is very limited, it is not clear that their application would have a significant impact on animal numbers.

The ICCVAM final BRD should discuss the practice of, and the U.S. Federal restrictions on, the reuse of rabbits in pyrogenicity testing.

A discussion on the ethical cost of conducting concurrent RPT testing should be added.

10.2 Requirement for the Use of Animals

Is the discussion of the use of cultured human cells and the need for volunteers for donations of peripheral blood used in the *in vitro* test methods appropriate and adequate?

No, the licensing arrangements and the maintenance of the MM6 cell line are unclear.

The discussion that reduction of the use of animals (i.e., rabbits) will be associated with the increased use of another animal (i.e., humans) is inadequate.

11.0 PRACTICAL CONSIDERATIONS

11.1 Transferability of the *In Vitro* Pyrogen Test Methods

Are the following aspects of *in vitro* test method transferability, including an explanation of how this compares to the transferability of the RPT, adequately described with regard to the:

11.1.1 Facilities and major fixed equipment needs?

Yes, either a sterile tissue culture facility or a laboratory animal facility is needed.

11.1.2 General availability of other necessary equipment and supplies?

Yes, equipment and supplies for both *in vitro* and *in vivo* studies are routinely available. In general, the skills and kits required are available in most diagnostic and testing facilities.

The availability (in ready to use kit form), the convenience, and the lower costs of the BET will mitigate against widespread use of the *in vitro* pyrogen tests that are far more work intensive (e.g., cytokine and endotoxin standard curves must be established, tests must be performed in quadruplicate, multiple donors are required), less convenient (as yet only one of the assays is available in kit form), and probably associated with higher costs.

11.1.3 Nature of the drug substance tested?

Yes, the drug substances are adequately described. The overall requirements for the assays are comparable with most other types of *in vitro* QC diagnostic assays.

11.2 Personnel Training Considerations

Are the following aspects of the *in vitro* test method training adequately considered? Is the explanation of how this compares to the level of training required to conduct the RPT adequate with respect to:

11.2.1 The required level of training and expertise needed to conduct the test method?

Yes, the individual technical steps and competencies are common to many other laboratory activities.

11.2.2 Any training requirements needed for personnel to demonstrate proficiency and any laboratory proficiency criteria that should be met?

The training required for adequate conduct of biological assays cannot be overestimated. Aseptic tissue culture techniques are essential, as is the ability to accurately serially dilute material. It is necessary to maintain the MM6 cell line and functional and non-activated monocytes obtained from whole blood. Activation can be caused by physical disruption or contaminants. Competency in each of these techniques should be demonstrated prior to allowing personnel to carry out these tests on medicinal products intended for human use or for certification. It should also be noted that the required expertise needed does not typically reside in the laboratories that conduct the test (i.e., RPT) targeted for replacement by the proposed *in vitro* tests.

11.3 Cost Considerations

Is the cost involved in conducting a study using the *in vitro* test method, as compared to the cost of conducting the RPT, adequately evaluated, and is this considered to be cost-effective compared to the *in vivo* method?

No, the direct and indirect costs of operating an animal facility that would be needed to house rabbits are incompletely stated. The *in vitro* pyrogen tests would seem to be considerably more cost effective than the RPT. It would be interesting to see pricing costs from contract research organizations for both classes of tests, mindful that cost considerations will impact on the level of use.

11.4 Time Considerations

Is the amount of time needed to conduct a study using the *in vitro* test method as compared to the time it takes to conduct the RPT adequately evaluated, and is the *in vitro* test method considered to be time-effective compared to the *in vivo* method?

The *in vitro* pyrogen test methods require two days to complete (twice as long as the BET and RPT under normal circumstances). Furthermore, the *in vitro* pyrogen test methods are dependent on the availability of donors or blood supplies, which might further restrict the frequency to which these tests can be performed.

12.0 RECOMMENDED ADDITIONAL REFERENCES

Are all relevant publications referenced in the ICCVAM draft BRD? If not, what additional references should be included?

The following references should be included:

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

Brunson KW, Watson DW. 1974. Pyrogenic specificity of Streptococcal exotoxins, Staphylococcal enterotoxin, and Gram-negative endotoxin. *Inf Immun* 10(2):347-351.

Burrell R. 1994. Human responses to bacterial endotoxin. *Circ Shock* 43(3):137-153.

Dinareello CA. 2004. Infection, fever, and exogenous and endogenous pyrogens: some concepts have changed. *J Endotoxin Res* 10(4):201-222.

Dixon WJ. 1950. Analysis of extreme values. *Ann Math Stat* 21(4):488-506.

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(1-2):165-177.

Grubbs FE. 1950. Sample criteria for testing outlying observations. *Ann Math Stat* 21(1):27-58.

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

Ishii KJ, Coban C, Akira S. 2005. Manifold mechanisms of toll-like receptor-ligand recognition. 2005. *J Clin Immunol* 25(6):511-521.

Ishii KJ, Akira S. Innate immune recognition of, and regulation by, DNA. 2006. *Trends Immunol* 27(11):525-532.

Marth E, Kleinhappl B. 2002. Albumin is a necessary stabilizer of TBE-vaccine to avoid fever in children after vaccination. *Vaccine* 20(3-4):532-537.

Martich GD, Boujoukos AJ, Suffredini AF. 1993. Response of man to endotoxin. *Immunobiol* 187(3-5):403-416.

Martin MA, Roberts S. 2006. An evaluation of bootstrap methods for outlier detection in least squares regression. *J Appl Stat* 33(7):703-720.

Martis L, Patel M, Giertych J, Mongoven J, Tammine M, et al. 2005. Aseptic peritonitis 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 vaccination. *J Inf Dis* 194(4):444-453.

Mullington J, Korth C, Hermann DM, Orth A, Galanos G, Holsboer F, Pollmächer T. 2000. Dose-dependent effects of endotoxin on human sleep. *Am J Physiol Regulatory Integrative Comp Physiol* 278(4):947-955.

Netea MG, Kallberg BJ, van der Meer JWM. 2000. Circulating cytokines as mediators of fever. *Clin Inf Dis* 31(Suppl 5):S178-S184.

Norata GD, Garlaschelli K, Ongari M, Raselli S, Gigore L, Benenuto F, Maggi FM, Catapano AL. 2005. Effect of Toll-like receptor 4 (TLR-4) variants on intima-media thickness and monocyte-derived macrophage response to LPS. *J Int Med* 258(1):21-27.

Rylander R. 2002. Endotoxin in the environment. *J Endo Res* 8(4):241-252.

Saper CB, Breder CD. 1994. The neurological basis of fever. *New England J Med* 330(26):1880-1886.

van Deventer SJH, Buller HR, ten Cate JW, Aarden LA, Hack EC, Sturk A. 1990. Experimental endotoxemia in humans: Analysis of cytokine release and coagulation, fibrinolytic, and complement pathways. *Blood* 76(12):2520-2526.

van Deventer SJH. 2000. Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med* 26(Suppl 1):S98-S102.

von Aulock S, Schroder NWJ, Gueninzius K, Traub S, Hoffmann 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 Inf Dis* 188(6):938-943.

13.0 Summary of Validation Status of the *In Vitro* Pyrogen Test Methods

Does the Panel agree that the applicable validation criteria have been adequately addressed in order to determine the usefulness and limitations of these *in vitro* test methods, to serve as a substitute for the RPT, for the identification of Gram-negative endotoxin on a case-by-case basis, subject to product specific validation?

Yes, the information is adequate with which to make an informed decision.

Does the Panel agree that the performance of these test methods in terms of their reliability and relevance support the proposed use of these test methods (i.e., the detection of Gram-negative endotoxin in materials that are currently tested in the RPT, subject to product specific validation to demonstrate equivalency to the RPT)?

No, refer to the reasons indicated in the responses to **Sections 1.0 to 12.0**.

Minority Opinion # 1 (Drs. Karen Brown, Albert Li, and Jon Richmond): The qualification in the above statement 'subject to product specific validation' should allow for a vote of yes.

Minority Opinion #2 (Dr. Peter Theran): It is not clear that the qualification included in the above statement would preclude the use of the *in vitro* test methods as replacements for the RPT in those circumstances where the BET is currently serving to replace the RPT.

B REVIEW OF ICCVAM DRAFT TEST METHOD RECOMMENDATIONS

1.0 Proposed Test Method Usefulness and Limitations

Does the Panel agree that the available data and demonstrated performance in terms of relevance (i.e., accuracy/concordance, sensitivity, specificity, positive, and negative predictivity, false positive and false negative rates) and reliability (i.e., intralaboratory repeatability and intra- and inter-laboratory reproducibility) support the ICCVAM draft recommendations for these *in vitro* test methods in terms of the proposed test method usefulness and limitations⁶?

The Panel does not agree with this statement for the following reasons:

The usefulness of these *in vitro* test methods for detecting Gram-negative endotoxin has not been properly assessed for concordance with the RPT or for relevance in comparison to the BET. Therefore, it is not possible to truly assess their usefulness and limitations. It is regrettable that their ability to detect non-endotoxins could not be demonstrated and validated due to the limitations of the validation and performance evaluation studies conducted.

Test materials in pure form may directly promote the formation and release of cytokines and thus, they may not be suited to evaluation by the *in vitro* methods.

As much effort as possible should be placed on truly demonstrating that these assays can be reliably used to detect non-endotoxin pyrogens in actual manufacturing settings for a wide variety of products. Otherwise, these assays have little advantage over the already established and widely used BET.

Mechanisms exist for test method developers to qualify their method on a case-by-case basis (i.e., 21 CFR 610.9). Therefore, the use of any recommended method should be subject to product specific validation to demonstrate equivalence as regulated by 21 CFR 610.9.

Minority Opinion (Dr. Peter Theran): This Panel has considered the failure to undertake additional RPTs a significant flaw in this validation RPT study and therefore proposed that, in the future, similar validation studies should use the RPT to provide concordance data. I have no objection to the performance of *in vitro* tests in parallel with rabbit tests, which are already scheduled to be performed, in order to achieve concordance data. But, it is my opinion, that any recommendation for *de-novo* parallel RPT should be accompanied by a statement, as follows: “The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis).” The inclusion of this statement would reinforce the importance of the 3R’s and would serve as a reminder of U.S. Federal law.

2.0 Proposed Test Method Standardized Protocols

Does the Panel agree that the available data and demonstrated performance in terms of relevance (i.e., accuracy/concordance, sensitivity, specificity, positive, and negative

⁶The ICCVAM draft recommendations state that there is sufficient information, based on validation studies with a limited number of pharmaceuticals, to substantiate the use of these test methods for the detection of pyrogenicity mediated by Gram-negative endotoxin in materials that are currently tested in the RPT, subject to product specific validation to demonstrate equivalency.

predictivity, false positive and false negative rates) and reliability (i.e., intralaboratory repeatability and intra- and inter-laboratory reproducibility) support the ICCVAM draft recommendations for these *in vitro* test methods in terms of the proposed test method standardized protocols?

The Panel agrees with this statement, provided that the following list of inadequacies⁷ within the proposed standardized protocols are fully addressed:

1. Donor-to-donor inflammatory response variation is problematic and therefore multiple donors should be used and the number used appropriately justified.
2. Restricting the protocols to a 'limits' test design, based on the intravenous fever threshold, for all test materials independent of administration route could be considered inappropriate. The threshold concentration for intrathecally-administered materials would be lower because of the reduced permissible endotoxin limit associated with these types of products. The use of a 'benchmark reference lot comparison' test design would alleviate the necessity to use such strict permissible 'limits'. Interestingly, in the two known examples where *in vitro* pyrogen test data have been considered by the FDA for release testing (cited in the ICCVAM draft BRD), 'benchmark reference lot comparison' test design protocols have been used.
3. The protocols do not include sufficient descriptions of donor selection criteria (e.g., volunteer or paid, recruitment process, etc.) and conditions for venipuncture (e.g., qualified phlebotomists, number and frequency of venipunctures, etc.). In practice, the requirement for blood donors to have taken no medication and the recommended CO₂ concentrations are more stringent than the provisions suggested in the draft recommendations.
4. The protocols are inconsistent in their acceptance criteria with respect to the number of blood donors. The IL-6 primary cell assays require four donors to be used for each test with acceptance criteria applied to each donor. The IL-1 assays show equal variability between donors, but do not require these acceptance criteria.
5. The suggested dilution scheme for the initial endotoxin standard and for the subsequent dilutions should not be recommended. The initial dilution of the endotoxin standard in two of the assays uses 20 µL into 1980 µL. The margin of error with a micropipette is such that even the smallest error at this initial dilution could affect the whole assay and is often the cause of a substantial proportion of assay variability. To reduce this potential problem, it is recommended that alternative dilution schemes be developed based on the accuracy of the micropipettes.
6. The use of in-house ELISA assays should not be recommended due to poor transferability and the potential for poor interlaboratory reproducibility associated with these assays.

⁷Based on the list of 20 separate inadequacies outlined in this report, three Panel members felt that this list would be better described as a list of "many and substantial" inadequacies.

7. The protocols should clearly specify the need for resonication and/or vortexing of any reference endotoxin solution prior to each use.
8. To adequately test for interference, spiked test samples containing endotoxin must be pre-incubated for a specified time prior to addition to the blood cells.
9. The following should be included in the revised protocols: a consistent number of donors to be used for all test methods; the acceptable range of cytokine response for each test method; the rules and the rationale for exclusion of low and/or high responders.
10. Three separate lots should be included in the pre-qualification of any test material, similar to the protocol used for the BET.
11. The protocol for the MM6 cell line describes procedures that would be used for collecting blood from donors. This point obviously is not required for this particular protocol.
12. The ECVAM protocols are very complete as to sources for all solutions, equipment, etc. required for testing. The ICCVAM protocols are less specific. More specific details on all test method protocol components should be included.
13. Intellectual property issues, as identified in the ICCVAM draft BRD, should be addressed in the ICCVAM protocols.
14. To prevent inactivation of LPS binding protein, it should be specified that FBS is heat inactivated at 55°C.
15. The symbols for correlation coefficient (r and r^2) are interchanged inappropriately.
16. On pages 14 and 21, lines 298 and 450 respectively in the PBMC/IL-6 protocol, the basis for the definition of low responders must be justified.
17. On page 15, line 325 in the PBMC/IL-6 protocol, the performance of monocyte counts using a hemocytometer is inaccurate compared to modern flow cytometric methods.
18. If a hemocytometer is used, specifications for the number of replicate determinations (e.g., at least duplicate), the minimum number of cells counted, and the magnification used must be stated.
19. On page 14, line 295 in the WB/IL-1 protocol, the statement "not taken any drug" is not sufficiently inclusive. This statement must also specify no over-the-counter medications or recreational drugs.
20. On page 20, line 411 in the WB/IL-6 protocol, the statement "If necessary, . . . endotoxin concentration can be modified" is insufficient. The modification of endotoxin concentration must be defined.

3.0 Proposed Test Method Performance Standards

Does the Panel agree that the available data and demonstrated performance in terms of relevance (i.e., accuracy/concordance, sensitivity, specificity, positive, and negative predictivity, false positive and false negative rates) and reliability (i.e., intralaboratory repeatability and intra- and inter-laboratory reproducibility) support the ICCVAM draft recommendations for these *in vitro* test methods in terms of the proposed test method performance standards?

The Panel does not agree with this statement, based on the inadequacies within the proposed performance standards outlined below.

Essential Test Method Components

1. A uniform CV criterion should be defined, which is adequately stringent. The reported range of 20% - 45% is inappropriate.
2. The number of individual blood donors used and/or the number of donors to be included in a pool of multiple donors should be defined, if deemed appropriate.
3. The stringency by which the endotoxin curves are validated should be defined (either by using a four-parameter logistic model or by checking that the OD concentration values ascend in a sigmoidal manner).
4. The use of CVs or any other measure of variability should be appropriately justified. If the data have been log-transformed, then CVs are not informative.
5. The following issues may overestimate the performance of the test methods:
 - a) The nature of the prediction model used for dichotomizing the results;
 - b) Experimental design and data analysis that might lead to overestimation of the sensitivity of the tests;
 - c) The nature and interpretation of the *in vivo* data used in the study;
 - d) The nature and cause of incorrect results and the lack of agreement within and between laboratories;
 - e) Whether the tests accurately estimate the actual concentration of the pyrogen and whether results met some pre-defined criteria of success.
6. In **Section 2.3.3.1**, a 'significant increase' is not defined. In **Section 2.3.6**, consideration should be given to adding Quality Assurance data and known biological properties under the 'test substances and control substances' heading.

Accuracy and Reliability Values

The demonstrated performance of certain aspects of several of the assays, particularly in terms of accuracy or relevance, yields some concern. Two of the assays have false positive rates in excess of 16%, which essentially means that approximately 1 in every 6 production lots could be unnecessarily prevented from being released, a rate unlikely to be accepted by manufacturers. A number of these performance characteristic issues can probably be explained by the fact that some of the spike concentrations used were very close to the 'limit' concentration criterion set.

If the intended use of the *in vitro* assays were only to detect Gram-negative endotoxin, it would seem very important to compare their performance in parallel validation studies that should include the BET. If the intended use of the *in vitro* methods is to evaluate substances containing endotoxin that are unable to be evaluated with the BET, then the parallel testing studies should include the RPT. This type of comparison has neither been made from the RPT (2-way parallel testing was also not performed on the endotoxin-spiked sample sets included in the validation studies cited in the ICCVAM draft BRD) nor the BET standpoint. The last thing one wants to recommend is an inferior performing assay to the one that is already established; similar or superior is fine.

Minimum List of Reference Substances

If the intent of the proposal was to replace the RPT with one or more of the *in vitro* test methods under consideration, then the *in vitro* test methods must be validated for all classes of substances (e.g., pharmaceuticals, biologicals, and implants) and medical devices that can be tested with the RPT. Validation of the *in vitro* test methods with pyrogens (e.g., LTA, components of viruses and fungi) other than endotoxin also needs to be conducted.

Minority Opinion (Dr. Peter Theran): This Panel has considered the failure to undertake additional RPTs a significant flaw in this validation study and therefore proposed that, in the future, similar validation studies should use the RPT to provide concordance data. I have no objection to the performance of *in vitro* tests in parallel with rabbit tests, which are already scheduled to be performed, in order to achieve concordance data. But, it is my opinion, that any recommendation for *de-novo* parallel RPT should be accompanied by a statement, as follows: “The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis).” The inclusion of this statement would reinforce the importance of the 3R’s and would serve as a reminder of U.S. Federal law.

4.0 Proposed Additional Studies

Does the Panel agree that the available data and demonstrated performance in terms of relevance (i.e., accuracy/concordance, sensitivity, specificity, positive, and negative predictivity, false positive and false negative rates) and reliability (i.e., intralaboratory repeatability and intra- and inter-laboratory reproducibility) support the ICCVAM draft recommendations for these *in vitro* test methods in terms of the proposed additional studies?

The Panel agrees that to better determine the potential of these test methods, the proposed additional studies should be performed using the test methods described in the ICCVAM draft BRD, taking into account the comments and recommendations detailed previously. The Panel recognizes that these test methods could be applicable to a wider range of pyrogens and test materials, provided that they are adequately validated for such uses. Wherever possible, either historical data from parallel studies conducted concurrently should be retrospectively evaluated or parallel testing should be conducted concurrently with RPT data generated for regulatory purposes.

The following additional recommendations are given:

1. A repository of test materials that have been identified clinically as pyrogenic would be invaluable for use in future validation studies and may allow such studies to be conducted without the further use of animals.
2. A 'limit' test design protocol and a 'benchmark reference lot comparison' test design protocol for each assay should be included.
3. Both endotoxin-spiked and non-endotoxin spiked samples should be included.
4. The non-endotoxin standards should be characterized as completely as possible prior to their use in any study and should satisfy the requirements set forth by ICCVAM for reference standards that are stated in the ICCVAM draft BRD.
5. The endotoxin-spike concentrations used for the performance assessment studies should not be so close to the positive test concentration limit, especially considering the relatively large enhancement and inhibition range permitted in the sample specific qualification investigations.
6. All aspects of the studies should be completely GLP compliant and importantly, the laboratories and results should be independently audited. This would include pre- and post-study audits of the laboratories.
7. The substances tested in the studies should also include products that have intrinsic pro-inflammatory properties.
8. A prospective study that includes 3-way parallel testing, with all of the *in vitro* assays (using both of the above mentioned protocol designs) being compared to the RPT and the BET, should be included to allow for complete concordance analysis and comparative performance assessment. These studies may be conducted with historical RPT data provided that the same substances (i.e., same lot) are tested in each method. Based on ethical and scientific rationale, the design of any side-by-side studies should be limited only to those that can gain more data than already available in the literature (i.e., data from parallel testing), most likely on the ability of the RPT and the *in vitro* pyrogen tests to detect non-endotoxin pyrogens.
9. Test substances that better represent all categories of sample types intended for testing by the methods (e.g., pharmaceuticals, biologicals, and medical devices) should be included. If relevant, extraction procedure protocols for the detection of pyrogens in medical device materials should be included.
10. The effects of direct administration of IL-1 and IL-6 to rabbits and the comparison of the resulting pyrogenic response with endotoxin-mediated pyrogenicity should be evaluated.
11. The correlation of IL-1 and IL-6 levels in the *in vitro* tests with levels produced in rabbits using similar doses of endotoxin should be evaluated.

The following statistical recommendations are noted:

1. For reliability assessments, formal hypothesis testing is essential with the alternative hypothesis being no different between groups.
2. For any additional studies, formal sample size calculations for equivalence testing should be made to determine that the required number of replicates needed to reject the null hypothesis (i.e., that there is a difference in reliability) at a given level of significance and power. If the study is not prospectively powered, the posterior power should be provided along with the observed significance level.
3. The proposed strategy for the Cryo WB/IL-1 test method is to retest if a test fails because of too much variability. The statistical properties of this multistage procedure should be characterized.

Minority Opinion (Dr. Peter Theran): This Panel has considered the failure to undertake additional RPTs a significant flaw in this validation study and therefore proposed that, in the future, similar validation studies should use the RPT to provide concordance data. I have no objection to the performance of *in vitro* tests in parallel with rabbit tests, which are already scheduled to be performed, in order to achieve concordance data. But, it is my opinion, that any recommendation for *de-novo* parallel RPT should be accompanied by a statement, as follows: “The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis).” The inclusion of this statement would reinforce the importance of the 3R’s and would serve as a reminder of U.S. Federal law.

C. OVERALL PEER REVIEW OUTCOMES

This international independent Peer Review Panel, consisting of 13 expert scientists from five different countries, provided comments and recommendations on the usefulness and limitations of five *in vitro* pyrogen test methods for the detection and quantification of Gram-negative endotoxin and on the ICCVAM draft test method recommendations on the use of these *in vitro* methods as partial replacements for the RPT. These remarks are summarized below.

- In general, the information presented in the ICCVAM draft BRD was sufficient for the purpose of determining the usefulness and limitations of these test methods for their proposed use and for adequately addressing the applicable validation criteria on the basis of the currently available evidence.
- The available data and demonstrated performance in terms of their reliability and relevance do not at this time support the ICCVAM draft proposed use for these test methods (i.e., as a partial substitute or replacement for the RPT, for the identification of Gram-negative endotoxin, on a case-by-case basis, subject to product specific validation). To better characterize the test methods and more clearly define their reliability and relevance, the Panel recommended that specific additional studies be performed using the ICCVAM proposed protocols, taking into account the Panel's comments and recommendations.
 - The lack of parallel testing in the *in vitro* tests and the RPT, and the resulting lack of concordance data, was considered to be a major limitation of the validation study design. For this reason, the Panel recommended that future studies include parallel testing. A minority opinion (Dr. Peter Theran) associated with parallel testing was expressed as follows: "The use of rabbits in new parallel tests for the validation of an *in-vitro* test should only be conducted after a vigorous search for a scientifically sound, non-animal alternative (i.e., the need for additional animal studies must be justified on a case-by-case basis)".
- The available data and demonstrated performance in terms of their reliability and relevance does not support the ICCVAM draft performance standards for these *in vitro* test methods for regulatory purposes.
- The information provided in the ICCVAM draft BRD supports the ICCVAM draft recommended protocols for these five *in vitro* test methods, providing that the list of inadequacies⁸ identified by the Panel with respect to reliability and relevance are fully addressed.
- These test methods could be applicable to a wider range of pyrogens and test materials, provided that they are adequately validated for such uses.
- It is critical to recognize, despite concerns about the performance of these five *in vitro* test methods, that a formal process exists for materials regulated under 21 CFR 610.9 to qualify these *in vitro* methods for the identification of Gram-negative endotoxin on a case-by-case basis, subject to product specific validation.

⁸Based on the list of 20 separate inadequacies outlined in this report, three Panel members felt that this list would be better described as a list of "many and substantial" inadequacies.