IN VITRO PYROGEN TEST USING HUMAN PBMCs (PBMC/IL-6)

Contents

1 RATIONALE FOR THE PROPOSED TEST METHOD 4

1.1 INTRODUCTION 4
1.2 REGULATORY RATIONALE AND APPLICABILITY 5
1.3 SCIENTIFIC BASIS FOR THE PROPOSED TEST METHOD 6

2 TEST METHOD PROTOCOL COMPONENTS 8

2.1 OVERVIEW OF TEST METHOD 8
2.2 RATIONAL FOR SELECTED TEST COMPONENTS 9
2.3 BASIS FOR SELECTION OF THIS TEST METHOD 13
2.4 PROPRIETARY COMPONENTS 13
2.5 REPLICATES 13
2.6 MODIFICATIONS APPLIED AFTER VALIDATION 14
2.7 DIFFERENCES WITH SIMILAR TEST METHODS 14

3 SUBSTANCES USED FOR VALIDATION 15

3.1 SELECTION OF SUBSTANCES USED 15
3.2 NUMBER OF SUBSTANCES 15
3.3 DESCRIPTION OF SUBSTANCES USED 16
3.4 SAMPLE CODING PROCEDURE 17
3.5 RECOMMENDED REFERENCE CHEMICALS 17

4 *IN VIVO* REFERENCE DATA ON ACCURACY 18

4.1 TEST PROTOCOL *IN VIVO* REFERENCE TEST METHOD. 18
4.2 ACCURACY 18
4.3 ORIGINAL RECORDS 19
4.4 QUALITY OF DATA 19
4.5 TOXICOLOGY 20
4.6 BACKGROUND ON ASSAY PERFORMANCE 20

5 TEST METHOD DATA AND RESULTS 21

5.1 TEST METHOD PROTOCOL 21
5.2 ACCURACY AND RELIABILITY 23
5.3 STATISTICS 29
5.4 TABULATED RESULTS 33
5.5 CODING OF DATA 35
5.6 CIRCUMSTANCES 35
5.7 OTHER DATA AVAILABLE 35

6 TEST METHOD ACCURACY 36

6.1 ACCURACY 36
6.2 CONCORDANCY TO *IN VIVO* REFERENCE METHOD 38
6.3 COMPARISON WITH REFERENCE METHODS 38
6.4 STRENGTH AND LIMITATIONS 39
6.5 DATA INTERPRETATION 39
6.6 COMPARISON TO OTHER METHODS 39

7 TEST METHOD RELIABILITY (REPEATABILITY/REPRODUCIBILITY) 40

7.1 SELECTION OF SUBSTANCES 40
7.2 RESULTS 40
7.3 HISTORICAL DATA 45
7.4 COMPARISON TO OTHER METHODS 45

8 TEST METHOD DATA QUALITY 46

8.1 CONFORMITY 46
8.2 AUDITS 46
8.3 DEVIATIONS 46
8.4 RAW DATA 46

9 OTHER SCIENTIFIC REPORTS AND REVIEWS 47

9.1 SUMMARY 47
9.2 DISCUSSION 48
9.3 RESULTS OF SIMILAR VALIDATED METHOD 48

10 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT) 51

10.1 DIMINISH ANIMAL USE 51
10.2 CONTINUATION OF ANIMAL USE 51

11 PRACTICAL CONSIDERATIONS 53

11.1 TRANSFERABILITY 53
11.2 TRAINING 53
11.3 COST CONSIDERATIONS 54
11.4 TIME CONSIDERATIONS 54

12 REFERENCES 55

13 CATCH -UP VALIDATION: HUMAN PBMC/IL-6 IN VITRO PYROGEN TEST WITH CRYOPRESERVED MONOCYTOID CELLS. 57

13.1 RATIONALE 57
13.2 TEST METHOD PROTOCOL COMPONENTS 57
13.3 SUBSTANCES USED FOR VALIDATION 57
13.4 TEST METHOD ACCURACY 57
13.5 TEST METHOD RELIABILITY (REPRODUCIBILITY) 60
13.6 CONCLUSION 61

14 SUPPORTING MATERIALS (APPENDICES) 62

14.1 STANDARD OPERATING PROCEDURE (SOP) OF THE PROPOSED METHOD 62
14.2 STANDARD OPERATING PROCEDURE (SOP) OF THE REFERENCE METHOD 62
14.3 PUBLICATIONS 62
14.4 ORIGINAL DATA 65
14.5 PERFORMANCE STANDARDS 65

Appendix A – Method protocol(s) and trial plan(s)

Appendix B – Hardcopies of relevant publications

Appendix C – List of abbreviations and definitions

1 Rationale for the Proposed Test Method

1.1 Introduction

1.1.1. *Describe the historical background for the proposed test method, from original concept to present. This should include the rationale for its development, an overview of prior development and validation activities, and, if applicable, the extent to which the proposed test method is mechanistically and functionally similar to a validated test method with established performance standards.*

Pyrogens, a chemically heterogeneous group of hyperthermia- or fever-inducing compounds, derive from bacteria, viruses, fungi or from the host himself, and provoke an immune response by producing endogenous pyrogens such as prostaglandins and the pro-inflammatory cytokines interleukin-1 (IL-1), interleukin-6 (IL-6) and tumour necrosis factor-α (TNF-α) (Dinarello, 1999). Depending on the type and amount of pyrogen challenge and the sensitivity of an individual, even life-threatening shock-like conditions can be provoked. To assure quality and safety of any pharmaceutical product for parenteral application in humans, pyrogen testing is therefore imperative.

Depending on the medicinal product, one of two animal-based pyrogen tests is currently prescribed by the respective Pharmacopoeias, i.e. the rabbit pyrogen test and the bacterial endotoxin test (BET). For the rabbit pyrogen test, sterile test substances are injected intravenously to rabbits and any rise in body temperature is assessed. This *in vivo* test detects various pyrogens but not alone the fact that large numbers of animals are required to identify a few batches of pyrogen-containing samples argues against its use. In the past two decades, the declared intention to refine, reduce and replace animal testing, has lowered rabbit pyrogen testing by 80% by allowing to use the BET as an *in vitro* alternative pyrogen test for certain medicinal products (Cooper et al, 1971).

Bacterial endotoxin, comprising largely lipopolysaccharide (LPS) from the cell wall of Gram-negative bacteria that stimulates monocytes/macrophages via interaction with CD14 and toll-like receptor 4 (TLR4) (Beutler and Rietschel, 2003), is the pyrogen of major concern to the pharmaceutical industry due to its ubiquitous sources, its stability and its high pyrogenicity. With the BET, endotoxin is detected by its capacity to coagulate the amoebocyte lysate from the haemolymph of the American horseshoe crab, *Limulus polyphemus*, a principle recognised some 40 years ago (Levin and Bang, 1964). In the US, *Limulus* crabs are generally released into nature after drawing about 20% up to 50% of their blood and therefore most of these animals survive. However, the procedure still causes mortality of about 30,000 horseshoe crabs per year, which adds to even more efficient threats of the horseshoe crab population like its use as bait for fisheries, habitat loss and pollution (http://www.horseshoecrab.org/).

As with the rabbit test the general problem of translation of the test results to the human fever reaction persists. Moreover, although being highly sensitive, the failure of the BET to detect non-endotoxin pyrogens as well as its susceptibility to interference by e.g. high protein or lipid levels of test substances or by glucans impedes full replacement of the rabbit pyrogen test. Hence, hundreds-of-thousands rabbits per year are still used for pyrogen testing.

A test system that combines the high sensitivity and *in vitro* performance of the BET test with the wide range of pyrogens detectable by the rabbit pyrogen test is therefore

required in order to close the current testing gap for pyrogens and to avoid animal-based tests. With this intention and due to improved understanding of the human fever reaction (Dinarello, 1999), test systems based on *in vitro* activation of human monocytes were developed. First efforts date back about 20 years, when peripheral blood mononuclear cells (PBMC) were used to detect endotoxin by monitoring the release of pyrogenic cytokines (Duff and Atkins, 1982; Dinarello et al, 1984). Meanwhile, a number of different test systems, using either whole blood, peripheral blood mononuclear cells (PBMCs) or the monocytoid cell lines MONO MAC 6 (MM6) or THP-1 as a source for human monocytes and various read-outs were established (Poole et al., 1988; Ziegler et al, 1988; Tsuchiya et al, 1980; Hartung and Wendel, 1996; Hartung et al, 2001; Poole et al, 2003). These test systems were validated with the aim of developing a tool for formal inclusion into Pharmacopoeias, an important basis for implementing novel alternative pyrogen tests for product-specific validation.

1.1.2 *Summarize and provide the results of any peer review conducted to date and summarize any ongoing or planned reviews.*
When the validation study was commenced, the proposed test method had not been subjected to any other peer review process. In the meantime, the PBMC/IL-6 test developed by Novartis and Baxter Healthcare has been subjected to a rigorous peer-review by the US FDA and approved as an end-product release test (New Drug Application Number 16-267/S-037 approved on April 24, 2002).
All of the five methods are currently under peer review of the ECVAM Scientific Advisory Committee.

1.1.3 *Clearly indicate any confidential information associated with the test method; however, the inclusion of confidential information is discouraged.*
This document does not contain any confidential information.

1.2 Regulatory rationale and applicability

1.2.1 *Describe the current regulatory testing requirement(s) for which the proposed test method is applicable.*
To assure quality and safety of pharmaceutical products for parenteral application in humans, pyrogen testing is imperative. Depending on the drug, one of two pyrogen tests is currently prescribed by the European Pharmacopoeia, i.e. the rabbit pyrogen test and the bacterial endotoxin test (BET), and other national and international guidelines.

1.2.2 *Describe the intended regulatory use(s) (e.g., screen, substitute, replacement, or adjunct) of the proposed test method and how it will be used to substitute, replace, or complement any existing regulatory testing requirement(s).*
Dependent on the product and the presence of relevant clinical data on unexpected pyrogenicity of clinical lots, the proposed test method may be an alternative method for pyrogen testing, thus substituting the rabbit pyrogen test or in rare cases the BET. In certain cases, the proposed test method may function as a supplementary test method to assess compliance to the licensing dossier.
In case the proposed test method is an alternative for pyrogenicity testing, a thorough cross-validation between the proposed test method and the original method for the

specific medicinal product is warranted. In case the proposed test method is an adjunctive test to screen for (unexpected) pyrogenic lots, alert and alarm limits may be established based on consistency of production lots or (preferably) based on actual clinical data.

1.2.3 *Where applicable, discuss the similarities and differences in the endpoint measured in the proposed test method and the currently used in vivo reference test method and, if appropriate, between the proposed test method and a comparable validated test method with established performance standards.*
The current *in vivo* method (rabbit test), as described in the pharmacopoeia, and the proposed *in vitro* test method each determine very different end-points, though the biochemical origins of the response are similar.
The *in vivo* method determines the total rise in body temperature (fever induction) of the animals subjected to the medicinal product, as a result of pyrogens present in the product.
The proposed PBMC/IL-6 test is an *in vitro* model for the fever response mechanism. It determines the release of cytokines by monocytoid cells into the culture medium upon the interaction of pyrogens and specific receptors on the monocytoid cells. It is these cytokines that trigger the fever response *in vivo*.
Main difference between the *in vivo* and *in vitro* method are that the latter is quantitative and uses cells of human origin, thus better reflecting the physiological situation.

1.2.4 *Describe how the proposed test method fits into the overall strategy of hazard or safety assessment. If a component of a tiered assessment process, indicate the weight that should be applied relative to other measures.*
The proposed PBMC/IL-6 test method may be applied for those medicinal products for which the rabbit test is the only or most reliable method for pyrogenicity testing, since a) the medicinal product is not compatible with the BET or b) the medicinal product contains pyrogens other than Gram-negative endotoxin.
Limit concentrations for pyrogens are established based on consistency lots or actual clinical data or, in the case of endotoxin the endotoxin limit concentration (ELC) as defined for many medicinal products.

1.3 Scientific basis for the proposed test method

1.3.1 *Describe the purpose and mechanistic basis of the proposed test method.*
The proposed *in vitro* method is intended to determine the presence of pyrogens in medicinal products for parenteral use. The proposed test method is an *in vitro* model of the human fever response. It determines the release of cytokines upon the interaction of pyrogens and specific Toll-like receptors on the monocytoid cells (Beutler and Rietschel, 2003). It is these cytokines that trigger the fever response *in vivo*.

1.3.2 *Describe what is known and not known about the similarities and differences of modes and mechanisms of action in the proposed test method as compared to the species of interest (e.g., humans for human health-related toxicity testing).*
An important feature of the proposed test method is that it is based upon the use of monocytoid cells of human origin. It therefore by definition resembles more closely the actual response of humans. The two other test methods make use of either crustaceans (BET) or rabbits, both species more or less distinct from the human species. The response

of humans, horseshoe crabs and rabbits toward Gram-negative endotoxin has been studied extensively and the methods appear equivalent for this particular pyrogen (Cooper et al 1971; Greisman and Hornick, 1969). However, there are documented cases of medicinal products and specified pyrogenic substances that yield false-positive or false-negative results in either test method. Since the proposed test method is based on human cells, it may therefore predict more accurately the pyrogenicity of such substances in humans.

1.3.3 *Describe the intended range of substances amenable to the proposed test method and/or the limits of the proposed test method according to chemical class or physicochemical factors.*

The proposed test method is intended for the assessment of pyrogens in all parenteral medicinal products for human use, chemical or biological and including raw materials, bulk ingredients and excipients. Use of the proposed test method in testing environmental samples or medicinal products is suggested and may be feasible, but substantiating data are as yet limited or absent.

2 Test Method Protocol Components

2.1 Overview of test method

Provide an overview of how the proposed test method is conducted. If appropriate, this would include the extent to which the protocol for the proposed test method adheres to established performance standards.

A highly detailed protocol describing the proposed test method (*Detailed protocol PBMC/IL-6: "In Vitro Pyrogen Test Using PBMC (SP+PB var. Novartis) 03 10 02; electronic file name: SOP-PBMC IL 6*) is attached in Appendix A of this Background Review Document (BRD).

The PBMC/IL-6 test method is an assay for the detection of pyrogenic contamination. The test protocol itself can be divided into the following two parts:

1. Incubation of the sample with peripheral blood mononuclear cells (PBMCs).
2. An enzyme linked immunoassay (ELISA) for the measurement of IL-6.

Ad 1.

Human whole blood is collected by venipuncture into heparinized tubes for blood sampling. The blood is diluted with PBS and the PBMCs are obtained by density gradient centrifugation (Lymphoprep). PBMCs from 4 donors are incubated overnight (16-24 hours) together with saline and the sample of interest in sterile and pyrogen-free cell culture plate.

Ad 2.

After mixture of the samples, the concentrations of the cytokine IL-6 in the PBMC-conditioned medium are quantified using a ELISA specific for IL-6 as follows. Aliquots are added to the wells of a microtiterplate coated with IL-6 specific monoclonal antibodies. An enzyme-conjugated polyclonal antibody against IL-6 is added. During a subsequent incubation, a sandwich complex consisting of two antibodies and the IL-6 is formed. Unbound material is removed in between incubation steps by a washing of the ELISA-plate.

A chromogenic substrate (3,3′,5,5′ -tetramethylbenzidine [TMB]) reactive with the enzyme label is added. Color development is terminated by adding a stopping solution. The resulting color, read at the appropriate wavelength, is directly related to the IL-6 concentration.

The IL-6 ELISA used throughout this study is an in-house ELISA, developed by Novartis, in which the IL-6 calibrant is calibrated against the International Standard (IS) for IL-6 (WHO code: 89/548).

The WHO-LPS standard (code 94/580, E.coli O113:H10:K-), was used throughout the validation. This standard is identical to USP Reference Standard Endotoxin (EC6).

There are several possibilities to estimate the pyrogenic contamination of the preparations under test: 1) A quantitative estimation can be achieved by the construction of a dose-response curve for endotoxin standard (e.g. 5.0, 2.5, 1.0, 0.5 and 0.25 EU/ml) versus

Optical Density (OD)-value of the IL-6 ELISA. The contamination of the preparations is expressed in Endotoxin Unit Equivalents. 2) A qualitative test can be achieved by the inclusion of an endotoxin threshold control (e.g. one fixed dilution of the standard curve) which allows for the classification in positive and negative samples (i.e. pyrogenic and non-pyrogenic samples). 3) A qualitative test can also be achieved by inclusion of an appropriate positive product control.
A detailed description of analysis methods used during the validation of the test-method can be found in section 5 of the current BRD.

2.2 Rational for selected test components

Provide a detailed description and rationale, if appropriate, for the following aspects of the proposed test method:
2.2.1 Materials, equipment, and supplies needed.
The materials, equipment and supplies used for the PBMC/IL-6 test method are laboratory items that will be already available in a routine QC laboratory. There is no need for sophisticated or dedicated laboratory equipment throughout the test.
For all steps in the procedure, excluding the ELISA procedure, the materials (e.g. tips, containers, solutions) which will be in close contact with samples and blood cells need to be sterile and pyrogen free. The materials, equipment and supplies are specified in the method protocol (Appendix A). It should be noted that equivalent devices may also be used and it is the user's responsibility to validate the equivalence.

Materials for part 1: PBMC-Incubation
Equipment
- Incubator (37°C, 5% CO_2 humidified air)
- Class 2 laminar flow sterile cabinet
- Centrifuge (suitable for 50 ml centrifuge tubes)
- Vortex

Consumables
- Lymphoprep (Nycomed, Oslo, Norway)
- Fragmin
- 30 ml syringe with luer lock adapter
- 19 mm, 21 gauge butterfly needle (or 40 mm, 21 gauge butterfly needle)
- 5, 10, 25 ml serological pipettes.
- Pipettes suitable for 50 or 100 µl
- 96-wells tissue culture plates
- WHO-LPS standard

Materials for part 2: ELISA procedure
Equipment
- Multichannel pipettor.
- Pipettes suitable for 50, 100 or 1000 µl.
- Microplate mixer
- Microplate washer
- Microplate reader capable of readings at the appropriate wavelength

- A software package facilitating data generation, analysis, reporting, and quality control

Consumables

- Graduated cylinder and plastic storage containers
- 96-wells microtiter plates
- Mouse monoclonal anti-IL-6 antibody from clone 16 (Novartis)
- Horseradish peroxydase conjugated sheep polyclonal anti-Il-6 antibody.
- Human Interleukin-6 standard
- Coating buffer, blocking buffer, dilution buffer, stopping solution and wash solution as detailed in the test method protocol.

The IL-6 ELISA used is an in-house assay developed in the Novartis laboratory and uses the WHO IL-6 international standard. Any commercially available IL-6 ELISA kit using the same standard or a standard calibrated versus it may be used (if validated for this in-vitro pyrogen test). Including the appropriate positive and negative controls in each run ensures the reliability and accuracy of the PBMC/IL-6 test method. As a positive control a specified amount of the Endotoxin Standard is used. The assay should be considered acceptable only if the criteria described in the method protocol are met. Also the criteria for allowed variability of replicates within an assay have to be met. The IL-6 standard curve is an additional control of the performance of the assay.

2.2.2 Dose-selection procedures, including the need for any dose range-finding studies or acute toxicity data prior to conducting a study, if applicable.

For every kind of test compound the interference with PBMC and the Il-6 ELISA is determined. For this purpose a preliminary "dose finding" test is conducted to establish a suitable dilution for every new test compound. For the validation study (as described in section 4 of this BRD), the tested products were diluted according to their known ELC, which was usually far beyond interfering concentrations. The ELC of the tested products or drugs were calculated according to the European Pharmacopoeia. If no endotoxin limit is defined it can be estimated by dividing 350 EU by the maximum hourly dose (example: the maximum hourly dose is 100 mg/patient, then the estimated endotoxin limit is 350/100=3.5EU/mg).

2.2.3 Endpoint(s) measured.

The proposed test method is an *in vitro* model of the fever response mechanism. It determines the release of interleukin-6 (IL-6) by monocytoid cells isolated from human blood. IL-6 is released into the culture medium upon the interaction of pyrogens and specific receptors on the monocytoid cells. The measured endpoint IL-6 is one of the cytokines that trigger the fever response *in vivo*.

2.2.4 Duration of exposure.

The PBMCs are exposed to possible pyrogenic components in samples at 37°C for 16-24 hours in an atmosphere of 5% CO_2 in humidified air. This conditioned mixture containing endogenous pyrogens released by the cells, is subsequently assayed in the IL-6 ELISA.

2.2.5 Known limits of use.
The PBMC/IL-6 test as described in the test method protocol is not a finalized test system for the testing of all medicinal products. The method may be applied only to preparations that have been validated with this method, i.e. shown not to interfere with the blood and the Il-6 readout system at a specified dilution of the preparation. A paragraph describing the interference testing is included in the method protocol (Appendix A). However, at this moment there are no medicinal products known that cannot be tested with the method.

2.2.6 Nature of the response assessed.
The proposed test method is an *in vitro* model of the fever response mechanism. Upon the interaction of exogenous pyrogens and specific receptors on the monocytoid cells endogenous pyrogens (e.g. interleukins, TNF-□ and prostaglandins) are produced. In the body the fever response is triggered by these endogenous pyrogens. Immunoreactive IL-6, the measured endpoint for the current method, is one of these endogenous pyrogens.

2.2.7 *Appropriate vehicle, positive, and negative controls and the basis for their selection.*
Throughout the development and validation phase the test compounds are diluted in 0.9% (w/v) clinical saline. This 0.9% clinical saline is considered an appropriate vehicle as no interference with active substances of a drug is to be expected.
In addition the test includes several controls.
A negative control: 0.9% clinical saline (sodium chloride)
A positive control: WHO-LPS 94/580, 0.5 EU/ml in clinical saline.
A negative product control (NPC): clean, released batch for each drug.
A positive product control (PPC): test item spike with WHO-LPS (code 94/580) at 0.5 EU/ml.
The positive and negative controls are the same in every assay and are needed to establish the sensitivity of the test system. In addition, a product-based set of controls is used to reveal product related interference.

2.2.8 Acceptable range of vehicle, positive and negative control responses and the basis for the acceptable ranges.
A PBMC/IL-6 assay is considered acceptable for further analysis if the following criteria are met: The ELISA is valid if the OD of the blank control is below 0.15 and the mathematical function (quadratic model) of the IL-6 standard curve produces an $r^2 > 0.95$. The reactions (in terms of OD) on the endotoxin concentrations give a sigmoidal ascending dose response.

Exclusion criteria for blood donors (low and high responders):
Blood donors are considered low responders if their mean OD value for the endotoxin reference standard concentration 1 EU/ml is below the mean OD value for 1000 pg/ml of IL-6. Blood donors are considered high responder if the mean OD value for the negative control above the mean OD value at 500 pg/ml of IL-6

If the test samples show an irregular response (e.g. high SD), the results of the blank and the standard endotoxin concentrations derived from the donor are checked. The donor is excluded if these results are inconsistent with the ordinarily expected results.

As regards the substances to be tested, for products with an established ELC, specified in EU/ml, the product is diluted to its maximum valid dilution (MVD). The negative product control should be negative at the MVD. The response to the positive product control should be between 50% and 200% of the response to the positive control, indicating a possible pyrogenicity can be detected using these conditions.

However, for the results described throughout this BRD the data were accepted and analyzed according to the procedures described in section 5.3 "Statistics". This procedure was chosen as it allowed for a harmonized analysis of comparable data which were obtained with different *in vitro* pyrogen tests (i.e. MM6/IL-6, WBT/IL-1, WBT/IL-6).

2.2.9 Nature of the data to be collected and the methods used for data collection.
The raw data collected are the read-outs (absorbance) of the IL-6 ELISA, measured by an automated laboratory ELISA-plate reader. The wavelength is dependent on the chromogenic substrate applied. In the current study TMB is used and the ELISA-plate is read at a wavelength of 450 nm using a 540 nm to 590 nm corrective filter. The values of the measurement with the corrective filter is subtracted from values measured with the 450 nm filter.

2.2.10 Type of media in which data are stored.
Data are stored in electronic files (windows98 compatible software) and as hard copy.

2.2.11 Measures of variability.
As part of the development of the PBMC/IL-6 test method the intralaboratory repeatability was assessed by independent and identical replicated measurement of the different concentrations of WHO-LPS. Furthermore, the limit of detection and its dependence from known but uncontrollable variables such us operator and donor were investigated. These variables and the inherent variation of biological systems make up to the total variation of the method.

2.2.12 Statistical or non-statistical methods used to analyze the resulting data, including methods to analyze for a dose-response relationship. Justify and describe the method(s) employed.
The first phase of the study was performed with PBMCs from 1 individual donor. The large scale blinded studies with pharmaceutical drugs were run with PBMCs from 4 individual donors in parallel. A standard, using the International Standard for Endotoxin (calibrated in EU) is included, ranging from 0.063 EU/ml up to 1.0 EU/ml.

2.2.13 Decision criteria and the basis for the prediction model used to classify a test chemical (e.g., positive, negative, or equivocal), as appropriate.
A prediction model (PM) was developed in order to classify substances as "pyrogenic for humans" or "non-pyrogenic for humans". To be able to define a dichotome result in the

alternative pyrogen test, a threshold pyrogen value of 0.5 EU/ml was chosen. This threshold value was based on historical data with rabbits (described in section 4.1). The suitability of the PM was assessed by testing substances which were artificially contaminated with endotoxin. (substances are described in section 3.2 and 3.3)
The statistical approach, including quality criteria, is detailed in section 5.3

2.2.14 Information and data that will be included in the study report and availability of standard forms for data collection and submission.
Raw data were collected using a standard form. These were submitted to the quality department of ECVAM.

2.3 Basis for selection of this test method

Explain the basis for selection of the test method system. If an animal model is being used, this should include the rationale for selecting the species, strain or stock, sex, acceptable age range, diet, and other applicable parameters.
In view of the shortcomings of the rabbit pyrogen test and the BET, *in vitro* pyrogen tests that utilize the exquisite sensitivity to exogenous pyrogen of monocytoid cells have been proposed. In such tests, products are incubated with human cell and the conditioned media assayed for pyrogenic cytokines (Duff & Atkins, 1982; Dinarello et al., 1984; Poole et al., 1988; Poole et al, 1989; Hansen & Christensen, 1990; Taktak et al., 1991; Bleeker et al., 1994).
The human PBMC assay was developed as a real *in vitro* alternative to the rabbit pyrogen test. The basic idea was to mimic the fever reaction in humans. In general, the detection of exogenous pyrogens (e.g. endotoxin) by monocytoid cells causes them to release endogenous pyrogens like IL-1β, IL-6 and TNFα. These cytokines affect the thermal regulation centre in the brain and increase the body temperature by changing its set point. In the past several test methods have been developed that use the sensitivity of human peripheral blood monocytes to exogenous pyrogens. In an attempt to increase the sensitivity of these tests the monocytes/leukocytes were isolated from whole blood, as has been done for the PBMC/IL-6 test. An overview of relevant literature can be found in section 9 of this BRD. Interleukin IL-6 is chosen as the readout because IL-6, unlike IL-1 and TNF, is secreted entirely into the cell-conditioned medium in large quantities, thereby permitting its complete estimation.

2.4 Proprietary components

If the test method employs proprietary components, describe what procedures are used to ensure their integrity (in terms of reliability and accuracy) from "lot-to-lot" and over time. Also describe procedures that the user may employ to verify the integrity of the proprietary components.
The test method does not employ proprietary components.

2.5 Replicates

Describe the basis for the number of replicate and repeat experiments; provide the rationale if experiments are not replicated or repeated.
All experiments with medicinal drugs are run using individual PBMCs, isolated from 4 individual donors.

During a prevalidation phase, the intralaboratory reproducibility as well as the interlaboratory reproducibility of the PBMC/IL-6 test method was established by applying repeated experiments (see section 7). As the test method reliability (repeatability /reproducibility) was shown to be satisfactory, it was feasible to establish the accuracy using pharmaceutical substances (detailed in table 3.3.1) by one test performed by three participating laboratories (see section 6).

2.6 Modifications applied after validation

Discuss the basis for any modifications to the proposed test method protocol that were made based on results from validation studies.

During the first part of the validation, which was restricted to the testing of LPS-spiked saline, only one donor was used. To reduce the variation the test was performed with PBMCs of 4 donors assayed separately.

The test can easily be adjusted to a quantitative assay as described in the detailed method protocol. However, the assay has now been validated as a qualitative assay, by means of the PM.

2.7 Differences with similar test methods

If applicable, discuss any differences between the protocol for the proposed test method and that for a comparable validated test method with established performance standards.

Not applicable.

3 Substances Used for Validation

3.1 Selection of substances used

Describe the rationale for the chemicals or products selected for use in the validation process. Include information on the suitability of the substances selected for testing, indicating any chemicals that were found to be unsuitable.

Selected test items were medicinal products available on the market. Released clinical batches were considered clean, i.e. containing no detectable pyrogens. To test the specificity, sensitivity and the reproducibility of the proposed test method, the products were spiked with pyrogen. For the present studies endotoxin (LPS) was selected as the model pyrogen, since it is well defined, standardized and readily available.

For the sensitivity and specificity the test items were assessed at their maximal valid dilution (MVD). The MVD is the quotient of the ELC and the detection limit. The European Pharmacopoeia prescribes for various types of parenterals the amount of endotoxin that is maximally allowed in a medicinal product, i.e. the ELC, taking into consideration the dose, the route of administration and the dosing regimen of the product. The aim of the study was to discriminate between negative and positive samples. Based on the selected pyrogen threshold value (see 4.1) and the PM applied, positive samples were defined as containing 0.5 EU/ml or more. Hence, to determine the MVD, the value of 0.5 EU/ml was defined as the detection limit.
Test items were assessed as such (negative product control), spiked with endotoxin at 0.5 IU/ml (positive product control) and after spiking with endotoxin at 5 levels (blinded samples). In addition a negative control (saline) and positive control (0.5 EU/ml in saline) were included to establish assay validity.

For reproducibility, the test items were tested, at a predefined dilution above the MVD, independently in 3 different laboratories, 3 times each. Based on the selected pyrogen threshold value (see 4.1) and the PM applied, positive samples were defined as containing 0.5 EU/ml or more. The test items were tested after spiking with endotoxin at four levels. For no other reasons but practical ones, i.e. availability of test materials, different test items were selected for this part of the validation study.

It was determined earlier whether candidate test items interfered with the outcome of the proposed test method. Interference was considered when the response of endotoxin in the diluted test item was below 50% or above 200% of the response of endotoxin in saline (spike-recovery). It was shown that none of the test items interfered with the assay at the selected dilutions (data not shown).

3.2 Number of substances

Discuss the rationale for the number of substances that were tested.

A total of 13 test items were selected for the validation study (see 3.3): 10 test items for determining sensitivity and specificity (table 3.3.1), 3 different test items for determining reproducibility (table 3.3.2). Test items and their spikes were appropriately blinded by ECVAM before distribution to the participating testing facilities.

For sensitivity and specificity, each test item was tested after spiking at its individual MVD. Hence they each came with their own specific set of 5 endotoxin spike solutions: 0.0, 0.25, 0.5, 0.5 and 1.0 EU/ml. Simple logistics limited the amount of test items for this part of the validation study to 10. Since test items were assessed with 5 different endotoxin levels at 3 independent test facilities, this yielded a total of 150 data points, biometrically considered to be sufficient for further analysis.

For reproducibility each test item was spiked at 4 different levels (0.0, 0.0, 0.5 and 1.0 EU/ml) and tested at specified dilutions, 3 times at 3 laboratories.

3.3 Description of substances used

Table 3.3.1: Test items (parenteral drugs) used for determining sensitivity and specificity

Drug	code	Source	Agent	Indication	MVD (-fold)
Glucose 5% (w/v)	GL	Eifel	Glucose	nutrition	70
Ethanol 13% (w/w)	ET	B.Braun	Ethanol	diluent	35
MCP®	ME	Hexal	Metoclopramid	antiemetic	350
Orasthin®	OR	Aventis	Oxytocin	initiation of delivery	700
Binotal®	BI	Aventis	Ampicillin	antibiotic	140
Fenistil®	FE	Novartis	Dimetindenmaleat	antiallergic	175
Sostril®	SO	GlaxoSmithKline	Ranitidine	antiacidic	140
Beloc®	BE	Astra Zeneca	Metoprolol tartrate	heart dysfunction	140
Drug A*	LO	-	0.9% NaCl	-	35
Drug B*	MO	-	0.9% NaCl	-	70

*Drugs A and B were included as saline controls using notional ELCs.
Negative control: 0.9% clinical stock saline solution.
Positive control: WHO-LPS 94/580 (*E. coli* 0113:H10:K-), 0.5 EU/ml in clinical stock saline.

Table 3.3.2: Test items (parenteral drugs) used for determining reproducibility.

Drug	Source	Agent	Indication
Gelafundin®	Braun melsungen	Gelatin	Transfusion
Jonosteril ®	Fresenius	Electrolytes	Infusion
Haemate ®	Aventis	Factor VIII	Hemophilia

Negative control: 0.9% clinical stock saline solution.
Positive control: WHO-LPS 94/580 (*E. coli* 0113:H10:K-), 0.5 EU/ml in clinical stock saline.

3.4 Sample coding procedure

Describe the coding procedures used in the validation studies.

All test items are registered medicinal products and were obtained from a pharmaceutical supplier. Test items and endotoxin spiking samples were prepared, blinded where appropriate and coded under GLP by personnel from ECVAM, Italy. These were then taken over by the Paul-Ehrlich Institute, Germany, for testing-allocation and shipment to each of the appropriate test facilities participating in the study.

For the sensitivity and specificity part of this study, test items and their respective spikes (5 per test item) were all blinded. For reproducibility, only the spikes (4) were blinded, the test items were not.

3.5 Recommended reference chemicals

For proposed test methods that are mechanistically and functionally similar to a validated test method with established performance standards, discuss the extent to which the recommended reference chemicals were tested in the proposed test method. In situations where a listed reference chemical was unavailable, the criteria used to select a replacement chemical should be described. To the extent possible, when compared to the original reference chemical, the replacement chemical should be from the same chemical/product class and produce similar effects in the in vivo reference test method. In addition, if applicable, the replacement chemical should have been tested in the mechanistically and functionally similar validated test method. If applicable, the rationale for adding additional chemicals and the adequacy of data from the in vivo reference test method or the species of interest should be provided.

The reference pyrogen material used was the international endotoxin standard WHO-LPS 94/580 (*E. coli* 0113:H10:K-). Where appropriate, the material was diluted in clinical saline solution (0.9%(w/v) sodium chloride). The saline was also used as negative control (blank).

4 *In vivo* Reference Data on Accuracy

4.1 Test protocol *in vivo* reference test method.

Provide a clear description of the protocol(s) used to generate data from the in vivo reference test method. If a specific guideline has been followed, it should be provided. Any deviations should be indicated, including the rationale for the deviation.

For ethical reasons, no rabbit pyrogen tests were performed for this study. However, Dr. U. Lüderitz-Püchel, Paul-Ehrlich Institute, Germany, kindly provided historical data, accumulated over several years, from 171 rabbits (Chinchilla Bastards). The respective Pharmacopoeia's do not prescribe a rabbit strain for the *in vivo* pyrogen test, but Chinchilla rabbits are reported as a relatively sensitive strain for pyrogen testing.

The rabbits were injected with endotoxin and their rise in body temperature over the next 180 minutes was recorded (figure 4.1.1). From these data it was established that 50% of the rabbits got fever when treated with 5 EU/Kg (Hoffmann et al, 2005a). Fever in rabbits is defined as a rise in body temperature over 0.55°C. On the basis of these historical animal data and corrected for the maximal volume allowed in rabbits, i.e. 10 mls per animal, a pyrogen threshold value of 0.5 EU/ml was defined for the PM in the proposed test method.

4.2 Accuracy

Provide the in vivo reference test method data used to assess the accuracy of the proposed test method. Individual human and/or animal reference test data, if available, should be provided. Provide the source of the reference data, including the literature citation for published data, or the laboratory study director and year generated for unpublished data.

As mentioned, no animal studies were done for ethical reasons. However, a theoretical assumption on specificity and sensitivity can be made (Hoffmann et al, 2005a). Taking the prevalence of the different final concentrations of LPS in the 5 spikes into account (1.0 EU/ml: 20%; 0.5 EU/ml: 40%; 0.25 EU/ml: 20% and 0.0 EU/ml: 20%) and calculating probabilities of misclassification using the chosen study design and defined threshold of pyrogenicity, i.e. 0.5 EU/ml, the theoretical sensitivity of the rabbit pyrogen test is 57.9% and the theoretical specificity of the rabbit pyrogen test is 88.3%.

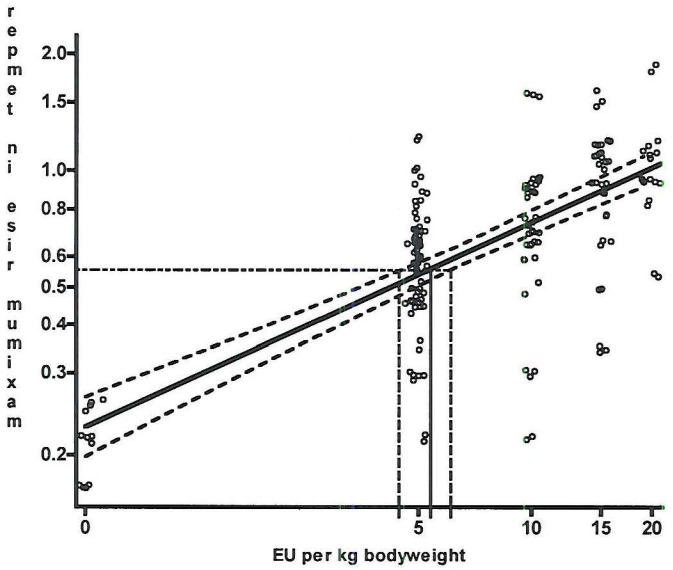

Figure 4.1.1 Dose-temperature of standard endotoxin applied to Chinchilla Bastards (n=171). Rabbits were treated with 1 ml saline containing 0, 5, 10, 15 and 20 EU of *E. coli* LPS (WHO-LPS 94/580 (E.coli O113:H10:K)) and their body temperature was measured over 180 min. Linear regression analysis was performed after logarithmic transformation of the data. Data are shown as dots to which a jitter-effect was applied in order to be able to distinguish congruent data. The full line depicts the linear regression whereas the dashed lines represent the 95%-confidence bounds. Furthermore, a horizontal line for a 0.55°C raise of temperature is added which is often defined as the rabbit threshold for fever. At the interception point of this line and the regression line 50% of the rabbits are to be expected to develop fever.

4.3 Original records

If not included in the submission, indicate if original records are available for the in vivo reference test method data.

The recognition of pyrogenic substances as bacterial by-products and the identification of a variety of pyrogenic agents enabled the development of a proper test to demonstrate non-pyrogenicity of the pharmaceutical product. As early as the 1920s, studies were done to select the most appropriate animal model. Results indicated that most mammals had a pyrogenic response, but only a few, including rabbits, dogs, cats, monkeys and horses showed a response similar to that in humans. For practical reasons, other species but rabbits and dogs were considered not practical. In 1942, Co Tui & Schrift described that rabbits are less thermo-stable as compared to dogs. Hence, rabbits are more suited for the purpose of testing for the absence of pyrogens, since a negative result is more significant.

4.4 Quality of data

Indicate the quality of the in vivo reference test method data, including the extent of GLP compliance and any use of coded chemicals.

All procedures employed in the study are GLP-concordant and quality assured by ECVAM's quality assurance officers.

4.5 Toxicology

Discuss the availability and use of relevant toxicity information from the species of interest (e.g., human studies and reported toxicity from accidental or occupational exposure for human health-related toxicity testing).

Over time, a number of studies were done to correlate the rabbit test to pyrogenic reactions in humans. A conclusive study by Greisman and Hornick, published in 1969, who compared three purified endotoxin preparations (*Salmonella typhosa*, *E. Coli* and *Pseudomonas*) in New Zealand rabbits and in male volunteers, showed that the induction of a threshold pyrogenic response, on a weight basis, was similar to rabbit and man. At higher doses, rabbits respond less severe as compared to man.

4.6 Background on assay performance

Discuss what is known or not known about the accuracy and reliability of the in vivo reference test method.

As mentioned, no animal studies were done for ethical reasons. However, a theoretical assumption on specificity and sensitivity can be made (Hoffmann et al, 2005a). Taking the prevalence of the different final concentrations of LPS in the 5 spikes into account (1.0 EU/ml: 20%; 0.5 EU/ml: 40%; 0.25 EU/ml: 20% and 0.0 EU/ml: 20%) and calculating probabilities of misclassification using the chosen study design and defined threshold of pyrogenicity, i.e. 0.5 EU/ml, the theoretical sensitivity of the rabbit pyrogen test is 57.9% and the theoretical specificity of the rabbit pyrogen test is 88.3%.

5 Test Method Data and Results

5.1 Test method protocol

Describe the proposed test method protocol used to generate each submitted set of data. Any differences from the proposed test method protocol should be described, and a rationale or explanation for the difference provided. Any protocol modifications made during the development process and their impact should be clearly stated for each data set.

The protocol for the PBMC/IL-6 test is provided Appendix A of this BRD. It includes the precise step-by-step description of the test method, including the listing of all the necessary reagents and laboratory procedures for generating data. For two steps during validation a part of the method protocol was adapted to contain a detailed description of the dilution of the samples and the spiking with WHO-LPS. The relevant part of the protocol is detailed in this section as well. The validity criteria and the detailed statistical analysis described in section 5.3 of this BRD were applied to analyse the data produced during validation.

To assess the reliability of the test method a series of experiments were conducted in the developing laboratory (DL). As a start, only blanks (saline, 0 EU/ml) and spikes of WHO-LPS in saline were tested. These experiments are summarised in table 5.1.1.

Table 5.1.1: summary of experiments with WHO-LPS in saline.

Experiment	Spikes (EU/ml) in saline	n (per spike)	Repetitions of experiment	N
1A	0; 0.25; 0.5	20	1	60
1B	0; 0.063; 0.125; 0.25; 0.5	12	1	60
2A	0; 0.5	8	3	48
2B	0; 0.063; 0.125; 0.25; 0.5	8	3	120
2C	0; 0.125; 0.25; 0.5	8	8	256

The collected data were used to answer questions regarding the nature of the distribution, the variance and its behaviour over the range of response in replicated measurements under identical conditions. In addition intralaboratory reproducibility was assessed by the maintenance of minimum criteria, e.g. a detection limit below an a-priori chosen positive control or a dose-dependent standard curve (table 5.1.1, experiment 1b). With the data of this experiment an assessment of the limit of detection of the test can be done by calculating the smallest spike, which can be discriminated from the blank.

The PBMC/IL-6 method was transferred from the DL to two other laboratories (denoted as naive laboratory 1 [NL1] and naive laboratory 2 [NL2]). A large scale dose response experiment was performed by all three laboratories. For this study 6 or 7 concentrations were tested in a dose response curve (typically 0; 0.125; 0.25; 0.5; 1; 2 EU/ml, at least 8 replicates) and all laboratories had to meet the validity criteria as laid down in the method protocol before the studies with medicinal substances were conducted.

The **intra- and interlaboratory reproducibility** was assessed by testing 3 different medicinal substances, Gelafundin, Jonosteril and Haemate (described in table 3.3.2). Test items and their spikes were appropriately blinded. Test items were tested, at a predefined dilution above the MVD, independently in 3 laboratories, 3 times each. Test items were tested after spiking with WHO-LPS at four different levels, the spikes were blinded and coded by QA ECVAM. In addition a negative control (saline) and positive control (0.5 EU/ml) in saline were included to establish assay validity.
Although this part of the study was designed for assessment of reproducibility, a preliminary estimate of the accuracy could be derived from the data. Applying the PM to the results and evaluating concordance in a two-by-two contingency table assessed accuracy.

To assess **accuracy** of the proposed test method 10 substances (listed in table 3.3.1), were spiked with five different concentrations of the WHO-LPS (one of which is negative). To permit the application of the chosen PM, each drug was diluted to its individual MVD, which has been calculated using the ELC to that drug (listed in table 3.3.1.) Each substance had their own specific set of endotoxin spike solution, ensuring that after spiking the undiluted drug, it contained 0.0; 0.25; 0.5; 0.5 and 1.0 EU/ml respectively when tested at its MVD. A detailed description of the sample preparation was supplied to the three independent laboratories (shown in table 5.1.2). To put more weight to the result of this part of the validation, the spikes were blinded and coded by QA ECVAM. The raw data of the PBMC/IL-6 assay are shown in paragraph 5.2. Accuracy was assessed by applying the PM to the results and evaluating the concordance in a two by two table. As intralaboratory reproducibility was (successfully) shown in previous experiments, only interlaboratory reproducibility was assessed in this phase.

Table 5.1.2: Sample preparation for the testing of 10 substances spiked with 5 different concentrations of WHO-LPS.

unblinded			*blinded*				
dilution of drug up to **MVD** ↓			spiking of **undiluted** drug: 0.5 ml each				
diluted drug	**NPC**	**PPC**	+ 23.3 µl of Spike 1	+ 23.3 µl of Spike 2	+ 23.3 µl of Spike 3	+ 23.3 µl of Spike 4	+ 23.3 µl of Spike 5
0.5 ml	+ 25 µl saline	+ 25 µl PPC-LPS-spike *					
		(final conc. = 50 pg/ml)	**dilution** to **MVD** ↓				
	test	test	test	test	test	test	test

* *PPC-LPS-spike contains 1050 pg/ml = 21fold 50 pg/ml*

NPC = Negative Product Control, **PPC** = Positive Product Control, **MVD** = Maximal Valid Dilution

5.2 Accuracy and reliability

Provide all data obtained to evaluate the accuracy and reliability of the proposed test method. This should include copies of original data from individual animals and/or individual samples, as well as derived data. The laboratory's summary judgment regarding the outcome of each test should be provided. The submission should include data (and explanations) from all studies, whether successful or not.

See figures 5.2.1, 5.2.2, 5.2.3, 5.2.4, 5.2.5 (A, B and C), 5.2.6 and 5.2.7 (A and B).

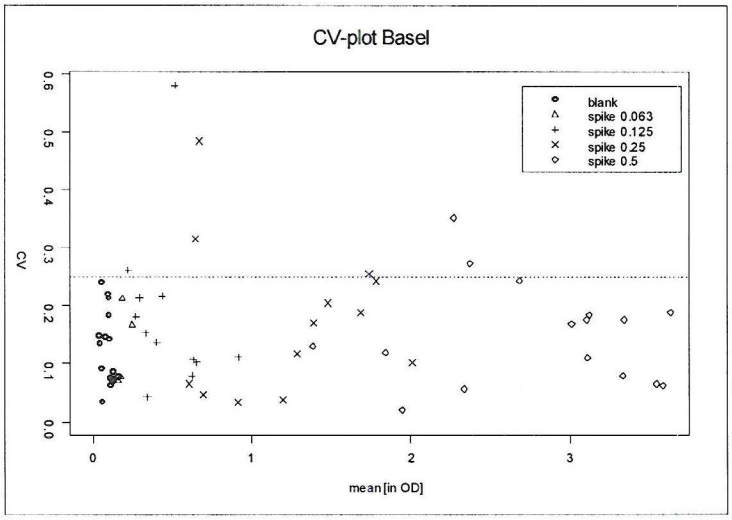

Figure. 5.2.1: Coefficient of variation (CV) of WHO-LPS spikes (4 replicates) relative to the mean OD (readout of the IL-6 ELISA).

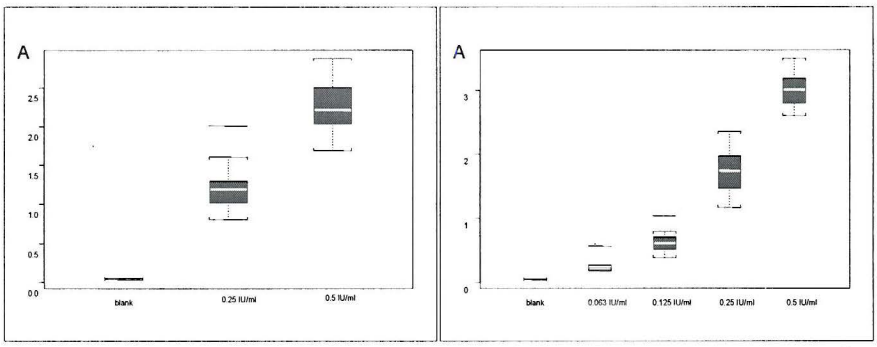

Figure. 5.2.2: Boxplots with absorbance (A) values of 20- replicates (left) or 12 replicates (right) of WHO-LPS spikes in saline at various concentrations. (readout of the IL-6 ELISA).

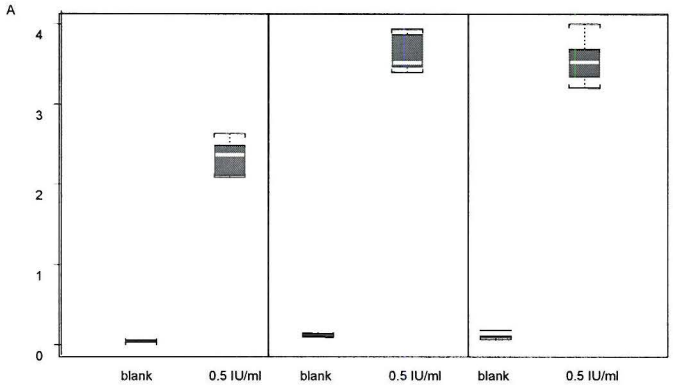

Figure. 5.2.3: Boxplots absorbance (A) values of the response of 3 different blood donations from one healthy volunteer on consecutive days with WHO-LPS (IU/ml) in saline at 0.0 IU/ml (Blank) or 0.5 IU/ml (readout of the IL-6 ELISA).

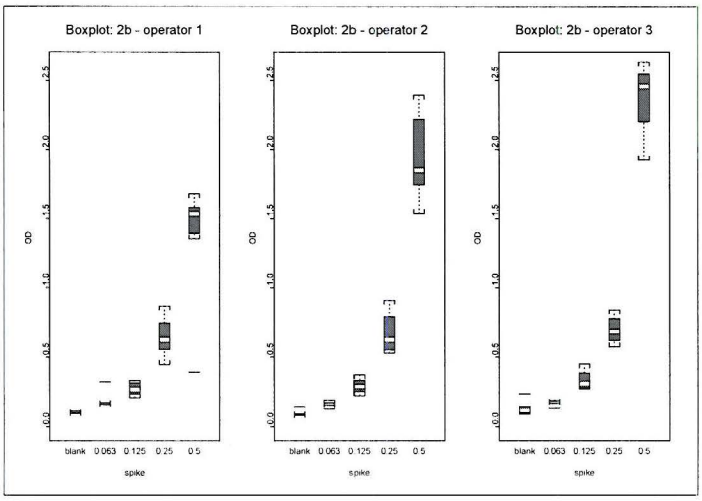

Figure. 5.2.4 A: Boxplots of absorbance (OD) values of WHO-LPS (IU/ml) in saline at 0.0 (Blank), 0.25 IU/ml (S0.25) or 0.5 IU/ml. (S0.5) with 3 different operators (readout of the IL-6 ELISA).

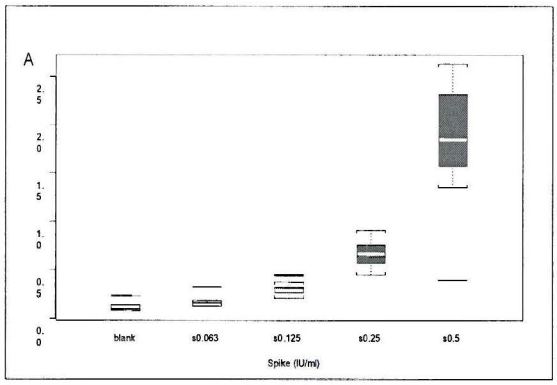

Figure. 5.2.4 B: Boxplot of absorbance (A) values of WHO-LPS (IU/ml) in saline at 0.0 (Blank), 0.25 IU/ml (S0.25) or 0.5 IU/ml. (S0.5). Combined data of Fig. 5.2.4.A (readout of the IL-6 ELISA).

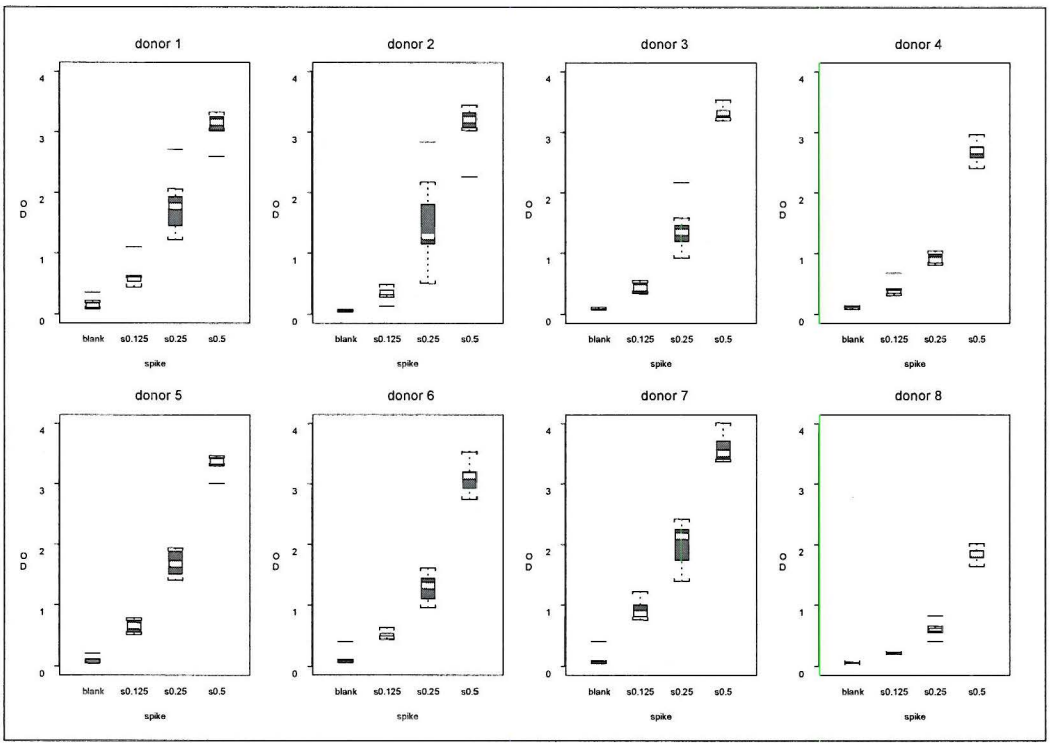

Figure. 5.2.5: Boxplots of absorbance values (OD) of the response of 8 individual donors to WHO-LPS (IU/ml) in saline at 0.0 IU/ml (blank), 0.125 IU/ml (s0.125), 0.25 IU/ml (s0.25) and 0.5 IU/ml (s0.5) (readout of the IL-6 ELISA).

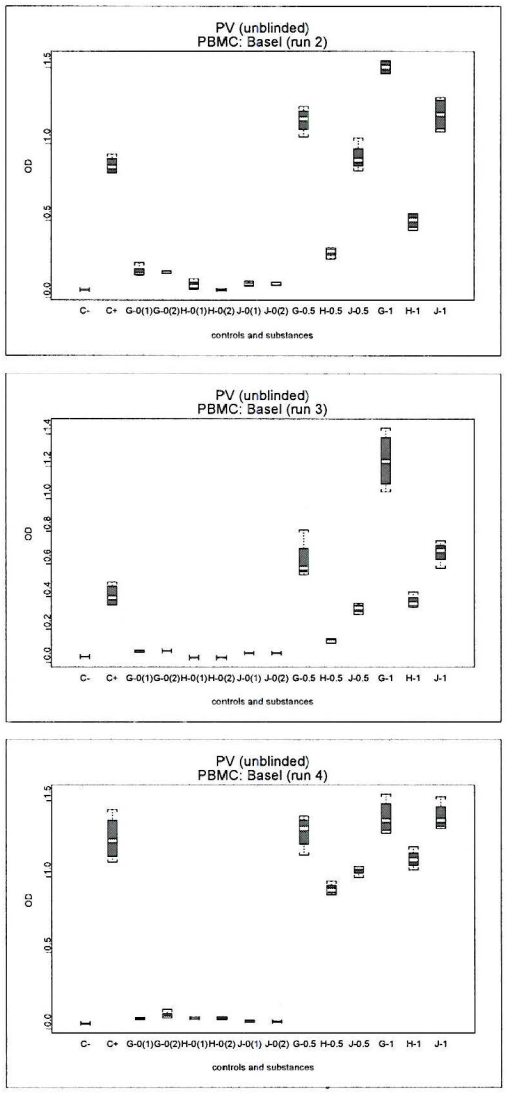

Figure. 5.2.6 A: Three different drugs were spiked with WHO-LPS at 0.0, 0.0, 0.5 and 1.0 IU/ml, respectively. Test was performed with known drugs and blinded spikes. Experiment was run 3 time independently at the Basel laboratory (readout of the IL-6 ELISA). The first run was invalidated for technical reasons (data not shown).
G = Gelafundin; J = Jonosteril; H = Heamate.
C- = negative control (saline); C+ = positive control (0.5 IU/ml in saline).

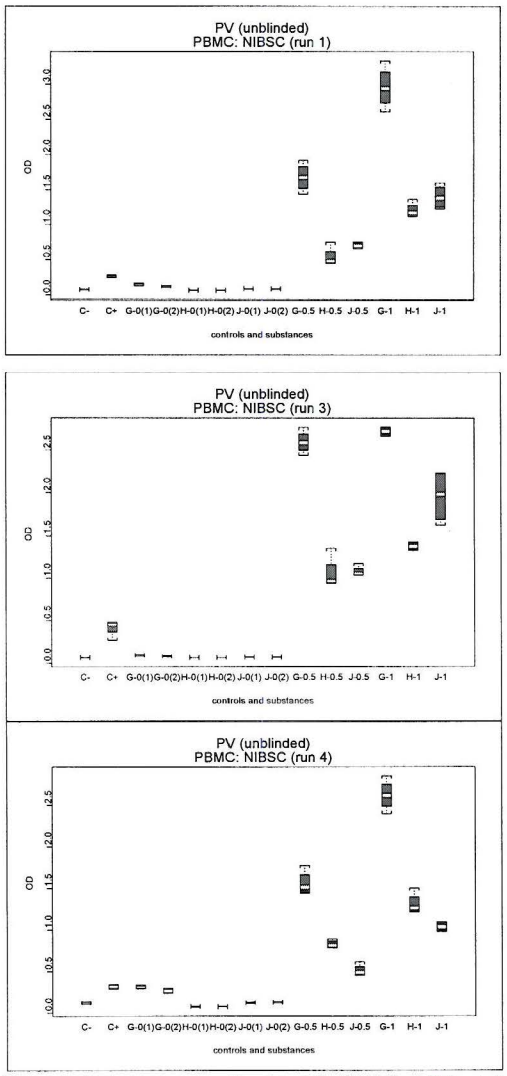

Figure. 5.2.6 B: Three different drugs were spiked with WHO-LPS at 0.0, 0.0, 0.5 and 1.0 IU/ml, respectively. The spikes were blinded. Experiment was run 3 time independently at the laboratory of the NIBSC (readout of the IL-6 ELISA).
G = Gelafundin; J = Jonosteril; H = Heamate.
C- = negative control (saline); C+ = positive control (0.5 IU/ml in saline).

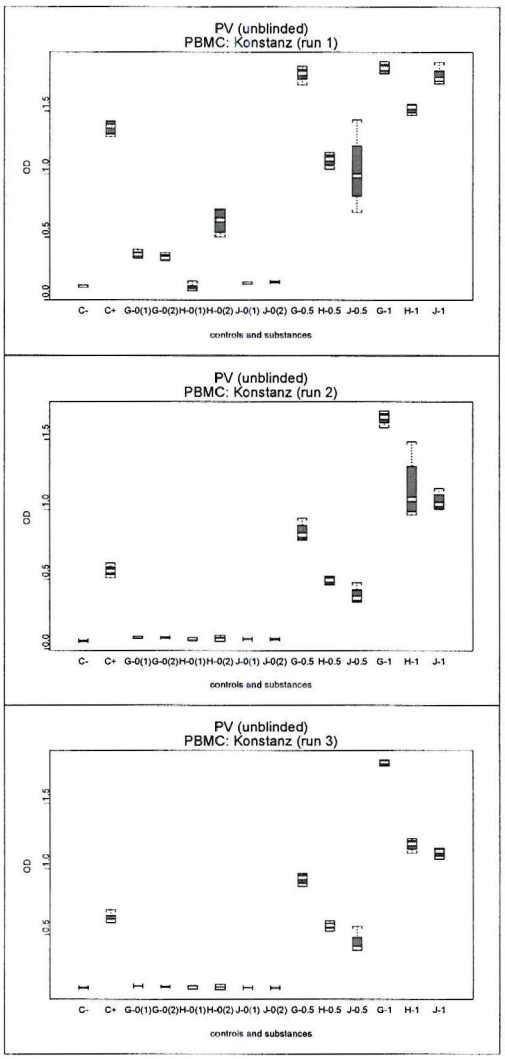

Figure. 5.2.6 C: Three different drugs were spiked with WHO-LPS at 0.0, 0.0, 0.5 and 1.0 IU/ml, respectively. The spikes were blinded. Experiment was run 3 time independently at the Konstanz laboratory (readout of the IL-6 ELISA).
G = Gelafundin; J = Jonosteril; H = Heamate.
C- = negative control (saline); C+ = positive control (0.5 IU/ml in saline).

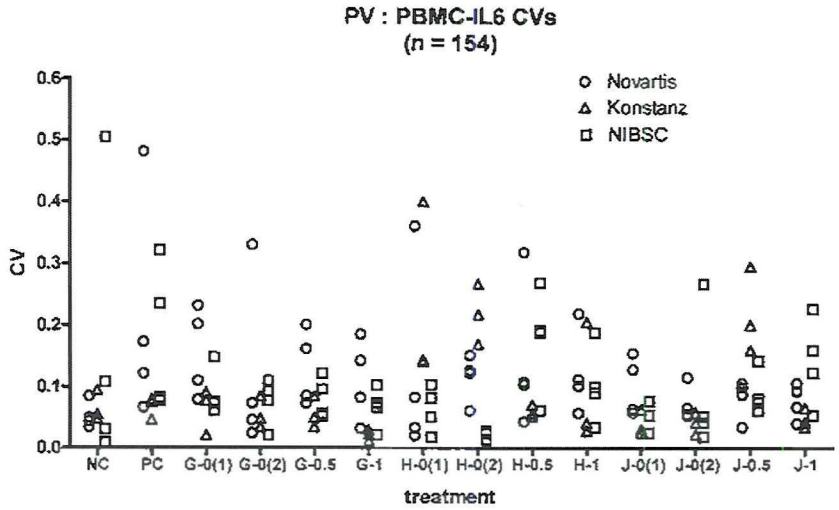

Figure. 5.2.7: Coefficient of variation (CV) of different WHO-LPS spikes (0.0, 0.0, 0.5 and 1.0 IU/ml, respectively).from the experiments as shown in fig. 5.2.6 A-C.
G = Gelafundin; J = Jonosteril; H = Heamate.
NC = negative control (saline); PC is positive control (0.5 IU/ml in saline).

5.3 Statistics

Describe the statistical approach used to evaluate the data resulting from studies conducted with the proposed test method.

A generally applicable analytical procedure was employed. This procedure includes a universal PM as well as quality criteria. First, a two-step procedure consisting of a variance-criterion and an outlier-test was applied. For this, the Dixon's test (Barnett and Lewis, 1984), which is USP approved, was chosen with the significance level of α=0.01 and applied to identify and eliminate aberrant data.

Next, the negative and the respective positive control are compared to ensure a suitable limit of detection. For this, a one-sided t-test with a significance level of α=0.01 is applied to the ln-transformed data to ensure that the response to the positive control is significantly larger than that of the respective negative control.

Finally, the samples are classified as either negative or positive by the outcome of a one-sided version of the t-test, which is based on the assigned pyrogen threshold value. The final results will be given in 2 x 2 contingency tables (table 5.3.1). These tables allow for estimation of accuracy (sensitivity and specificity) and reproducibility of the proposed test method.

Table 5.3.1: 2x2 contingency table.

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

Accuracy:

The most important statistical tool to determine accuracy (specificity and sensitivity) is the so-called PM (Hothorn, 1995). In general, it is a statistical model, which classifies a given drug by an objective diagnostic or deciding rule. The objective of a dichotomous result requires a clear cut PM, which assigns a drug in one of the two classes "pyrogenic for humans" and "non-pyrogenic for humans". Since a threshold pyrogen value will be used, a one-sided test is appropriate for the task. Because the data are normalised by a ln-transformation, a t-test is chosen. Although the variances over the range of concentration converge by the transformation, the assumptions of equal variances do generally not hold true, because it depends on additional covariates. Therefore, the one sided Welch-t-test (Snedecor and Cochran, 1989) is applied. Due to the safety aspect of the basic problem, the hypotheses of the test are

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

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

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

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

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

$$S_ij = 1, \text{ else},$$

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

Finally, the classifications of the drugs will be summarised in 2x2 contingency table (table 3). From these tables, estimates of the sensitivity (S_E), i.e. the probability of correctly classified positive drugs and specificity (S_P), i.e. the probability of correctly classified negative drugs, will be obtained by the respective proportions. Where

$$S_E = a / (a + c) * 100\%$$
and
$$S_P = d / (b + d) * 100\%.$$

Furthermore, these estimates will be accompanied by confidence intervals, which will be calculated by the Pearson-Clopper method [15]. For example, let $\hat{p}_{SE}$ denote the proportion, namely the sensitivity, under investigation. Then the confidence interval to a niveau α is calculated as

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

where $F_{...}$ denotes the respective quantile of the F-distribution and $n_{1.}$ is the sample size of the positive drugs and a the number of correctly classified drugs.

By contaminating the drugs artificially and by defining a threshold value, which is assumed to be appropriate, the class of a drug is determined beforehand. The versions of drugs, which are effectively contaminated, but below the threshold dose, are considered to be negative, respectively safe, because their contamination is not crucial for humans in terms of ELC.

Reproducibility:

The analysis of the intra- and interlaboratory reproducibility was assessed from the three identical and independent runs conducted in each of 3 laboratories. The comparison of

the three runs was carried out blindly such that the testing facility did not know the true classification of the sample, either pyrogenic or non-pyrogenic. By this procedure only the randomness and reproducibility of the methods was assessed and not systematic errors, which may have arisen from other sources, e.g. logistics or preparation of the samples. This means that if a sample was (mis)classified in all three runs the result is reproducible regardless of the (mis)classification of the sample. Therefore, a measure of similarity, i.e. complete simple matching with equal weights, was preferred to the coefficient of correlation for 2x2 contingency tables.
The study was designed as follows: each laboratory had to conduct three independent runs with the same 12 samples (3 test items with 4 blinded spikes each) and two controls, i.e. saline as a negative control (C-) and a 0.5 EU/ml LPS-spike in saline as a positive controls (C+). The samples were derived from the three substances Gelafundine, Haemate and Jonosteril. Per run, each substance was blindly spiked twice with saline, once with 0.5 EU/ml LPS and once with 1 EU/ml LPS, which resulted in a balanced design with regard to positive and negative samples, i.e. samples expected to be pyrogenic and non-pyrogenic, respectively.

The three independent runs per testing facility provide the information on which the assessment of the intralaboratory reproducibility is based. The combined results of the three runs per testing facility were used to determine interlaboratory reproducibility. The correlation of the prediction (in terms if the Bravais-Pearson coefficient of correlation) between all runs is calculated, independent of whether that classification is true or false. A BP-correlation of 1 is calculated, if two runs gave exactly the same predictions for the twelve substances. If one run gives adverse classifications for all substances than the other, the correlation is –1. As these calculations do not need information of the true status of a sample, they were carried out blinded.

5.4 Tabulated results

Provide a summary, in graphic or tabular form, of the results.
See tables 5.4.1 and 5.4.2.

Table 5.4.1: Results of testing 3 substances 3 times by 3 laboratories. Classifications after applying the PM (compare to fig. 5.2.5)

Sample	DL (Novartis, Basel)			NL 1 (Konstanz)			NL 2 (NIBSC)		
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
G-0 (1)	0	0	0	0	0	0	0	0	1
G-0 (2)	0	0	0	0	0	0	0	0	0
H-0 (1)	0	0	0	0	0	0	0	0	0
H-0 (2)	0	0	0	0	0	0	0	0	0
J-0 (1)	0	0	0	0	0	0	0	0	0
J-0 (2)	0	0	0	0	0	0	0	0	0
G - 0.5	1	1	1	1	1	1	1	1	1
H - 0.5	0	0	0	1	1	1	1	1	1
J - 0.5	1	1	1	0	0	0	1	1	1
G - 1	1	1	1	1	1	1	1	1	1
H - 1	1	0	1	1	1	1	1	1	1
J - 1	1	1	1	1	1	1	1	1	1

"0"denotes "non-pyrogenic"; "1" denotes "pyrogenic".

Table 5.4.2: Results of the validation study of 10 drugs, spiked with WHO-LPS at 0.0, 0.25, 0.5, 0.5 and 1.0 IU/ml, respectively and tested in 3 different laboratories. Samples and spikes were blinded. Classifications after applying the PM (compare to fig. 5.2.7).

drug (code)	spike EU/ml	"truth"	results Basel	Konstanz	NIBSC
Beloc (BE)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	0	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Binotal (BI)	0.00	0	0	0	0
	0.25	0	0	1	0
	0.50	1	1	1	0
	0.50	1	1	1	1
	1.00	1	1	1	1
Ethanol 13% (ET)	0.00	0	0	0	0
	0.25	0	0	1	0
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Fenistil (FE)	0.00	0	0	0	0
	0.25	0	0	1	0
	0.50	1	1	1	1

drug (code)	spike EU/ml	"truth"	results Basel	Konstanz	NIBSC
	0.50	1	1	1	1
	1.00	1	1	1	1
Glucose 5% (GL)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
"Drug A" 0.9% NaCl (LO)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	0	1	1
	0.50	1	0	1	1
	1.00	1	1	1	1
MCP (ME)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	1	1	0
	0.50	1	1	1	1
	1.00	1	1	1	1
"Drug B" 0.9% NaCl (MO)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	0	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Orasthin (OR)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Sostril (SO)	0.00	0	0	0	0
	0.25	0	0	0	0
	0.50	1	0	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1

"0"denotes "non-pyrogenic"; "1" denotes "pyrogenic"; NA is not assessed.

5.5 Coding of data

For each set of data, indicate whether coded chemicals were tested, whether experiments were conducted without knowledge of the chemicals being tested, and the extent to which experiments followed GLP guidelines.

Blinding of drugs and/or spikes is indicated with the data.

5.6 Circumstances

Indicate 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. A coded designation for each laboratory is acceptable.

In each part of the study, all samples are derived from one (clinical) lot.

5.7 Other data available

Indicate the availability of any data not submitted for external audit, if requested.

All relevant data were submitted with the present BRD.

6 Test Method Accuracy

6.1 Accuracy

Describe the accuracy (e.g., concordance, sensitivity, specificity, positive and negative predictivity, false positive and negative rates) of the proposed test method compared with the reference test method. Explain how discordant results in the same or multiple laboratories from the proposed test were considered when calculating accuracy.

Test method accuracy was assessed in two large scale experiments performed with the drugs outlined in table 3.3.1 and table 3.3.2 in section 3 respectively. As described before one experiment was performed in an early stage of the study with 3 different drugs, tested 3 times and the other final experiment all drugs were tested once in the three participating laboratories. From the first experiment a preliminary estimate of sensitivity and specificity can be figure out, whereas the second is regarded as the established accuracy for the PBMC/IL-6 assay.

6.1.1 Preliminary estimate of the accuracy of the PBMC/IL-6 test. the study a different concept for interference testing was used. The developing laboratories determined for each drug (outlined in table 3.3.2, section 3.3) the smallest dilution within the MVD that showed no interference or an acceptable degree of interference with the spike recovery. In general the lowest dilution of the sample allowing for a 50-200% spike recovery was chosen. In addition, the positive control (PC) set at 0.5 EU/ml saline was used as the classification threshold. The laboratory procedure as described in the protocol was maintained throughout the study. Although it was realized there were some drawbacks to the concept for interference testing and applying the PC as a threshold, this small scale study allows for a preliminary estimate of the accuracy of the PBMC/IL-6 method.

It has to be noted that this part of the study was designed to provide an estimate of the intra- and interlaboratory reproducibility. Therefore it will also be discussed in detail in section 7 (Test Method Reliability).

According to the PM applied during an early phase of the study the outcome (positive/negative) is related to the positive control (PC=0.5 EU/ml). If absorbance of sample > absorbance of PC, then the sample is classified as positive. If absorbance of sample < PC, then the sample is classified as negative. While performing the experiments during this phase it was realized that there is a flaw between interference testing and the PM. After the interference testing, the lowest dilution of the sample allowing for a 50-200% spike recovery was chosen. It is obvious that even with a flawless repeatability of the assay; a spike recovery between 50%-100% would be classified as negative according to the preliminary PM. In addition, due to unforeseen problems with the preparations of the spike, the recovery of the spikes was far below 100%. (This is outside the scope of the study and will not be discussed). As a consequence of the employed preliminary setup of the study the sensitivity will be underestimated, and the specificity will be overestimated.

In short, three test substances were spiked with WHO-LPS at four levels (0, 0, 0.5 and 1.0 EU/ml respectively), which resulted in 12 samples with a balanced design with regard to positive and negative samples (i.e. samples expected to be pyrogenic and non-pyrogenic respectively. These 12 samples were three times tested in three laboratories. In total there were 108 classifications from 12 samples in 3 runs and in 3 laboratories (3x3x12=108). Results are described in detail in section 7. A 2x2 contingency table was constructed (table 6.1.1), from which the estimates of sensitivity and specificity can easily be derived.

Table 6.1.1: 2x2 contingency table. The PM applied to a preliminary study.

	True status of samples +	-	**Total**
PM +	47	1	**48**
-	7	53	**60**
Total	**54**	**54**	**108**

The specifications of specificity and sensitivity described in section 5.3 were applied to these results.
The specificity (Sp) of the PBMC/IL-6 assay is 98.1% (53/(1+53)*100%), 95% confidence interval [0.901; 0.999]. The sensitivity (Se) equals 87% (47/(47+7)*100%), 95% confidence interval [0.751; 0.946]. As outlined previously the specificity is overestimated and the sensitivity is underestimated as a result of the design of this part of the study.

6.1.2 Test method accuracy of the proposed PBMC/IL-6 method. To assess accuracy of the proposed method, 10 substances (listed in table 3.1.1, section 3) were spiked with five different concentrations of the WHO-LPS (one of which is negative). Thus, in total, 50 samples have been tested in each laboratory.
To permit the application of the chosen PM, each drug was diluted to its individual MVD, which has been calculated using the ELC to that drug (listed in section 3). Lesser dilutions were tested by the DL, and showed no interference. Therefore interference was not expected at the individual MVD. Each substance had their own specific set of endotoxin spike solution, ensuring that after spiking the undiluted drug, it contained 0.0; 0.25; 0.5; 0.5 and 1.0 EU/ml respectively when tested at its MVD. A detailed description of the sample preparation was supplied to the three independent laboratories (shown in table 5.1.1 for convenience). To put more weight to the result of this part of the validation, the spikes were blinded and coded by QA ECVAM. The raw data and the graphical presentation of these raw data are shown in the section 5 (table 5.4.2).
Accuracy was assessed by applying the PM to the results (summarized in table 5.3.2) and evaluating the concordance in this section in a two by two contingency table (table 6.1.2). As described above 10 substances, spiked with 5 different WHO-LPS concentrations were tested in three laboratories and consequently a maximum of 150 data were available for analysis.

As intralaboratory reproducibility was successfully shown in previous experiments (analyzed in section 7), only one run performed in each laboratory was considered sufficient.

Table 6.1.2: 2x2 contingency table. PM applied to the PBMC/IL-6 test result of 10 different substances assessed in three different laboratories.

		True status of samples +	-	**Total**
PM	+	83	3	**86**
	-	7	57	**64**
	Total	**90**	**60**	**150**

All the 150 available sets of replicates met the quality criteria as described in section 5.3. The specificity and sensitivity of the PBMC/IL-6 method was estimated as described in section 5.3.

The specificity of the PBMC/IL-6 assay is 95.0% (57/(3+57)*100%), 95% confidence interval [0.861;0.990]. The sensitivity equals 92.2% (83/(83+7) *100%), 95% confidence interval [0846;0.968]. (See table 6.1.3). The specificity varied from 85% up to 100% within the three laboratories, and the sensitivity varied from 83.3% up to 100%.

Table 6.1.3: Specificity and sensitivity of the PBMC/IL-6 assay

	N total	**N correctly identified**	**proportion**	**95% CI lower limit**	**95% CI upper limit**
Specificity (Sp)	60	57	95.0%	86.1%	99.0%
Sensitivity (Se)	90	83	92.2%	84.6%	96.8%

6.2 Concordancy to *in vivo* reference method

Discuss results that are discordant with results from the in vivo reference method.
Not applicable.

6.3 Comparison with reference methods

Discuss the accuracy of the proposed test method compared to data or recognized toxicity from the species of interest (e.g., humans for human health-related toxicity testing), where such data or toxicity classification are available. This is essential when the method is measuring or predicting an endpoint for which there is no preexisting method. In instances where the proposed test method was discordant from the in vivo reference test method, describe the frequency of correct predictions of each test method compared to recognized toxicity information from the species of interest.
Not applicable.

6.4 Strength and limitations

State the strengths and limitations of the proposed test method, including those applicable to specific chemical classes or physical-chemical properties.

It appears the proposed test is applicable to most classes of medicinal products, at least those that are non- or low-toxic to cells *in vitro*. In addition, the test may be employed to assess pyrogenicity of various medical devices, such as (biological) bovine collagen bone implants.

6.5 Data interpretation

Describe the salient issues of data interpretation, including why specific parameters were selected for inclusion.

No issues.

6.6 Comparison to other methods

In cases where the proposed test method is mechanistically and functionally similar to a validated test method with established performance standards, the results obtained with both test methods should be compared with each other and with the in vivo reference test method and/or toxicity information from the species of interest.

Not applicable.

7 Test Method Reliability (Repeatability/Reproducibility)

7.1 Selection of substances

Discuss the selection rationale for the substances used to evaluate the reliability (intra-laboratory repeatability and intra- and inter-laboratory reproducibility) of the proposed test method as well as the extent to which the chosen set of substances represents the range of possible test outcomes.

The rationale for the selection of the substances is described in section 3.3. In short: for the present studies endotoxin (WHO-LPS) was selected as the model pyrogen, since it is well defined biological standard and readily available. Selected test substances were medicinal products available on the market. These batches are released by the manufacturers and comply with the Marketing Authorisation file and European Pharmacopoeia. Therefore these batches are considered to contain no *detectable* pyrogens. To test the method reliability the medical products were spiked with endotoxin.

7.2 Results

Provide analyses and conclusions reached regarding the repeatability and reproducibility of the proposed test method. Acceptable methods of analyses might include those described in ASTM E691-92 (13) or by coefficient of variation analysis.

In an early phase of the study, the intralaboratory repeatability and reproducibility of the test method was assessed in a series of experiments conducted in the DL). Series of blanks (saline, 0 EU/ml) and spikes of WHO-LPS in saline were tested. These experiments (1A, 1B, 2A, 2B and 2C) are summarized in table 7.2.1.

Table 7.2.1: Summary of experiments with WHO-LPS in saline.

Experiment	Spikes (EU/ml) in saline	n (per spike)	Repetitions of experiment	N
1A	0; 0.25; 0.5	20	1	60
1B	0; 0.063; 0.125; 0.25; 0.5	12	1	60
2A	0; 0.5	8	3	48
2B	0; 0.063; 0.125; 0.25; 0.5	8	3	120
2C	0; 0.125; 0.25; 0.5	8	8	256

The data were used to answer questions regarding the nature of the distribution, the variance and its behavior over the range of response in replicated measurements under identical conditions. In addition, reliability of the test method was assessed by the maintenance of minimum criteria, e.g. a detection limit below an a-priori chosen positive control or a dose-dependent standard curve (table 7.2.1, experiment 1B). With the data of this experiment an assessment of the limit of detection of the test can be done by calculating the smallest spike, which can be discriminated from the blank.

The second group of experiments was meant to analyze the variation in detail. For this purpose the major sources variation were assessed separately, i.e. behavior of a donor in time (experiment 2A), operator (exp. 2B) and different donors (exp.2C). A total of 554 observations were collected and analyzed.

First the shape of the distribution at a spike was assessed (not shown). Most of the data showed normal-distribution.
Based on the experience that there is a monotone increasing relationship between the mean-responses and the variation (empirical variance or standard deviation), the analysis focuses on the CV. Outliers were removed with the Grubbs-test (α =5%). The CV should be distributed symmetric around a constant factor, if the mean-variance relationship is linear. A plot of all CVs against their corresponding means is shown in figure 5.2.1. From the figure it is clear that at this stage of the study, the CV for most sets of replicates is lower than CV 0.25, only seven out of the 61 sets showed a CV above 0.25. This higher variation originates from only one of the experiments (experiment 2C), with 5 of the 7 higher values observed for donor 1 and donor 2.

The outliers were identified on the assumption of normally distributed data as well as a log-normal distribution. At this point the Grubbs-test was chosen and the kind of outlier (lower or upper) and the significance level α (5% and 1% significance level) were recorded. Altogether there were 19 outliers identified out of 544 observations, which is a proportion of only 3.5%. Most of the outliers originate from blanks, six out of the seven were upper outliers. This might be explained by small inhomogeneities, which will have a major impact. In addition, the raw data (plate-readouts) showed no obvious edge-effects or trends.

The results of test 1A (figure 5.2.2) show a low variation between the 20 replicates of each spike. The figure shows one outlier, which could detected with the Grubbs-test. The spike of 0.25 EU/ml was easily detected. Test 1B showed a dose response curve with 12 replicates for each spike concentration. Two outliers could be identified. It is obvious that that even a 0.0625 EU/ml of LPS standard can be detected with the PBMC/IL6 assay.

Test 2A (figure 5.2.3) was included to assess the behavior of a donor in time. Blood samples were collected at day 1, day 36 and day 41 respectively. Data are presented in figure 5.2.3. Unfortunately, the interpretation of the result is impaired because the first experiment was executed by another operator than the last two experiments. It is not clear whether the lower spike -response of the first day is due to the operator or to the day. However, in general the results of the three days are comparable and the 0.5 EU/ml spike can be detected beyond doubt.

Five LPS spikes were tested by three operators in parallel while conducting experiment 2B. The results, presented in boxplots in figure 5.2.4, are similar. Only the level of the readout of the 0.5 EU/ml spike differ. Again the 0.063 EU/ml spike was detected by every operator. It can be concluded that the operator does not have an important impact on the performance of the assay.

The last experiment 2C was designed to show the robustness of the assay with respect to different donors. Therefore 8 donors were involved and for each donor eight replicates of each of the spikes (0; 0.125; 0.25; 0.5 EU/ml) were generated. Figure 5.2.5. shows that the level of the OD-values differ from donor to donor due to their individual sensitivity to

LPS. But every donor reacts to the lowest spike tested, the 0.125 EU/ml-spike. As mentioned before a few very large variances, are noticeable for donor 1 and 2. Regarding the overall low variances, this is considered an incident. The donor will have an impact on the results of an assay in terms of reproducibility over donors, because the data differ stronly between donors. But in terms of a qualitative decision all donor behaved the same. To minimise the influence of different donors, acceptance criteria have already been included in the protocol for PBMC/IL-6 assay.

In conclusion:
The data showed a very stable statistical properties over all the experiments. The shape of replicates, the numbers of outliers and the variation reveal a general structure. Nevertheless the variation can be quite high. Even if the outliers are removed, 25% of the CVs are larger than 0.2 and 10% are over 0.25. This is an acceptable amount of variation, but this might be higher if the number of replicates is reduced to four.
The sensitivity of the assay is very high. Even the spike of 0.063 EU/ml can be detected with low errors-rates (false-positive, false-negative). The two covariates "day" and "operator" have just a small impact on the result of the assay. The influence of the covariate "donor" cannot be neglected with regard to reproducibility of the OD. Nevertheless, for every donor the spikes could be discriminated from each other. Therefore the intralaboratory repeatability is considered satisfactory. The 3-4% percentage outliers, as determined by the Grubbs test is considered acceptable.

Intra- and interlaboratory reproducibility.
After transfer of the PBMC/IL-6 assays to two other laboratories, a dose response experiments was performed by all three laboratories. For this study 7 concentrations were tested in a dose response curve (0, 0.125, 0.25, 0.5, 1, 2, 4 EU/ml, at least 8 replicates). A participating laboratory qualified for taking part in the next part of the study by producing a dose response curve, with a limit of detection of at least 0.5 EU/ml and a CV < 0.4 (data not shown).

Testing 3 different medicinal substances, Gelafundin, Jonosteril and Haemate (described in table 3.3.2, section 3.3.) assessed the intra- and interlaboratory reproducibility. Test substances and their spikes were appropriately blinded. Test substances were tested, at a predefined dilution above the MVD, independently in 3 different laboratories, 3 times each. The three test substances were spiked with WHO-LPS at four levels (0, 0, 0.5 and 1.0 EU/ml respectively), which resulted in 12 samples with a balanced design with regard to positive and negative samples (i.e. samples expected to be pyrogenic and non-pyrogenic respectively. In addition a negative control (saline) and positive control (0.5 EU/ml in saline) were included to establish assay validity. To avoid interference, the DL performed interference testing in terms of the BET, i.e. 50-200% spike recovery, and decided on the dilution of the test substances. Dilutions chosen for Gelafundine, Haemate, Jonosteril were 1:4, 1:2 and 1:1 respectively. The results are graphically presented using the absorbance values of the three runs (shown in section 5, fig. 5.2.5).

From the experiment with LPS-WHO only it was concluded that CV for the PBMC/IL-6 assay is < 0.4, which is acceptable. It was envisaged that the CV was likely to be higher when testing different substances (different matrices) and was assessed for the current set of data. A plot of all CVs for all sets of 4 replicates of a spiked drug is shown in figure 5.2.7. From the figure it is clear that the CV for a set of 4 replicates of one spike concentration is usually below 0.45, which is considered acceptable for a biological assay. Only one set of data showed an exceptional high (CV>1.1) which is probably due to a pipetting error. For the remainder of the studies the CV criteria applied as validity criteria of the PBMC/IL-6 assays was arbitrarily set at CV<0.45.

The analysis of the intralaboratory reproducibility was assessed from the three identical and independent runs conducted in each laboratory. The comparison of the three runs was carried out blindly such that the laboratory did not know the true classification of the sample (either pyrogenic or non-pyrogenic). By this procedure only the randomness and reproducibility of the methods was assessed and not systematic errors, which may have arisen from other sources, e.g. logistics or preparation of the samples. This means that if a sample was misclassified in all three runs the result is 100% intralaboratory reproducible (regardless of the misclassification of the sample).

According to the preliminary PM applied during this phase of the study the outcome (positive/negative) was related to the positive control (PC=0.5 EU/ml). If absorbance of sample > absorbance of PC, then sample is classified as being positive. If absorbance of sample < PC, sample is classified as negative (positive/pyrogenic = 1, negative/non-pyrogenic = 0).
During this phase it was realized that there is a flaw between interference testing and the PM. After the interference testing, the lowest dilution of the sample allowing 50-200% spike recovery was chosen. It is obvious that even with a flawless repeatability of the assay, a spike recovery between 50%-100% would be classified as negative according to the preliminary PM.

From the three independent runs summarized in table 5.4.1, the intralaboratory reproducibility can be calculated for the separate laboratories (table 7.2.2). For these calculations there is no need for information of the true status of the sample. A minimum criterion for the establishment of an assay is that experiments carried out with the same samples should result in a high concordance of classifications.

Table 7.2.2: Intralaboratory reproducibility, assessed by correlation between different runs. Result of testing 3 substances 3 times by 3 laboratories.

	DL (Basel)	**NL1 (Konstanz)**	**NL2 (NIBSC)**
Run 1 - Run 2	92% (11/12)	100% (12/12)	100% (12/12)
Run 1 - Run 3	100% (12/12)	100% (12/12)	92% (11/12)
Run 2 - Run 3	92% (9/12)	100% (12/12)	92% (11/12)
Mean	94%	100%	94%
Proportion showing the same result in 3 runs	92% (11/12)	100% (12/12)	94% (11/12)

Each of the assays performed by the laboratories fulfilled the sensitivity criterion, i.e. the assays showed a significant difference between C- and C+. All results could be included in the analysis. From table 7.2.2 it can be read that the between runs reproducibility ranges from 92 to 100%. The mean intralaboratory reproducibility is very good (94 to 100%) for all three participating laboratories.

The interlaboratory reproducibility of the PBMC/IL-6 method was assessed in a similar manner to the intralaboratory reproducibility. A summarizing method to combine the three runs per laboratory is considered not appropriate, because it would mask misclassification. Therefore each run of one laboratory was compared with all runs of another laboratory. This results optimally in 108 comparisons between the data sets of two laboratories. The measure of similarity is then the proportion of equally classified samples. These proportions are summarized in table 7.2.3, show that there is a good interlaboratory reproducibility varying from 81% to 89% (overall mean: 85%).

Table 7.2.3: Interlaboratory reproducibility: assessed by inter-laboratory correlations. Result of testing 3 substances 3 times by 3 laboratories.

Laboratories	Interlaboratory Reproducibility	Number of equal predictions
DL – NL1	81%	87 / 108
DL – NL2	86%	93 / 108
NL1 – NL 2	89%	96 / 108
Mean	85%	

DL = Basel; NL1 = Konstanz; NL2 = NIBSC

Also from the result of the large-scale study (testing 10 substances spiked with 5 separate spikes), the interlaboratory reproducibility can be estimated (table 7.2.4). The reproducibility varied from 84% to 90% between two laboratories. All three laboratories found the same result for 40 out of 50 samples (equals 80%).

Table 7.2.4: Interlaboratory reproducibility: Assessed by testing of 10 substances, spiked 5 times. One run of 50 samples by three different laboratories.

Laboratories	Interlaboratory Reproducibility	Number of equal predictions
DL - NL1	84%	42 / 50
DL - NL2	86%	43 / 50
NL1 – NL2	90%	45 / 50
Mean	87%	
same result in all labs	80%	40 / 50

DL =Basel; NL1 = Konstanz; NL2 = NIBSC

Conclusion: It is shown that the mean intralaboratory reproducibility, assessed by the proportion of equally classified samples between different runs varies from 94% to 100%

between the three participating laboratories. The interlaboratory reproducibility between two laboratories varied from 81% to 89% in one large scale blinded experiment and from 84% to 90% in the other large scale blinded experiment. All three participating laboratories predicted the same in 80% of the measurements. It has to be noted that a substantial part of the samples was 0.5 EU/ml and therefore close to the defined pyrogenicity-threshold of the PBMC/IL-6 assay

7.3 Historical data

Summarize historical positive and negative control data, including number of experiments, measures of central tendency, and variability.
Not applicable.

7.4 Comparison to other methods

In cases where the proposed test method is mechanistically and functionally similar to a validated test method with established performance standards, the reliability of the two test methods should be compared and any differences discussed.
Not applicable.

8 Test Method Data Quality

8.1 Conformity

State the extent of adherence to national and international GLP guidelines (7-12) for all submitted data, including that for the proposed test method, the in vivo reference test method, and if applicable, a comparable validated test method. Information regarding the use of coded chemicals and coded testing should be included.

The studies were done in concordance with the guidelines for GLP. Written protocols and approved standard operating procedures were followed during the entire course of the study. Deviations were recorded and, where appropriate, approved in amendments. All data are stored and archived. As mentioned, samples were appropriately blinded.

8.2 Audits

Summarize the results of any data quality audits, if conducted.

No audits were conducted.

8.3 Deviations

Discuss the impact of deviations from GLP guidelines or any noncompliance detected in the data quality audits.

Not applicable.

8.4 Raw data

Address the availability of laboratory notebooks or other records for an independent audit. Unpublished data should be supported by laboratory notebooks.

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

9 Other Scientific Reports and Reviews

9.1 Summary

Summarize all available and relevant data from other published or unpublished studies conducted using the proposed test method.

Relevant data obtained with the proposed method are described in a number of published studies and reports, which are added in the Appendix B.

An *in vitro* monocyte activation test that detected pro-inflammatory and pyrogenic contaminants, was first applied some 15 years ago (Poole et al., 1988). A number of variants of the original test system have since been described, although the underlying principle of each variant remains the same. The test preparation is cultured with monocytes, either as peripheral blood mononuclear cells, PBMCs, diluted whole blood or cells of a monocytoid cell line such as MONO MAC-6 (MM6). Contaminants in the test sample activate CD14/TLR receptors, which stimulate the release of an endogenous pyrogenic cytokine from the monocytes (Poole and Gaines Das, 2001).

Early studies report on opimization of the test method, e.g. improving the lower limit of detection, incubation times and cytokine readout, using model pyrogens such as LPS. Limited information is available on the actual testing of medicinal products.

Most interestingly, Taktak et al (1991) described several batches of a medicinal product (serum albumin) that caused adverse (pyrogenic) reactions in recipients. These lots were not detected by either BET or rabbit test but only by the *in vitro* monocytoid cell test.

In a study using whole blood and monocytoid cell lines as the sources of monocytoid cells (Nakagawa et al., 2002) it was reported that the structurally diverse pyrogens endotoxin, peptidoglycan, Staphylococcus aureus, Cowan 1, Curdlan and poly(I.C) all stimulated the release of cytokines.

The cytokine readout included tumour necrosis factor (TNF), interleukin-1 (IL-1), and IL-6 (reviewed by Poole and Gaines Das, 2001 and Poole et al., 2003). Other cytokines, e.g. IL-8, are also produced in large quantities in response to pyrogenic contaminants but their roles in fever are less well studied. The preferred readout is usually IL-6 because IL-6, unlike IL-1 and TNF, is secreted entirely into the cell-conditioned medium in large quantities, thereby permitting its complete estimation (Poole et al, 1988; Poole et al., 1989, Taktak et al., 1991).

Also, certain pro-inflammatory bacterial components stimulate the production of IL-6 but not TNF and IL-1 (Reddi et al., 1996), and IL-6 induction via Toll-like (pyrogen) receptors rapidly follows the recognition of microbial products (Pasare and Medzhitov, 2003).

It is stressed throughout these studies using whole blood and PBMCs that only healthy donors not taking any medication must be used for testing pharmaceuticals. Various drugs such as cortisone suppress interleukin release and would therefore exclude a blood donor from the experiment. An infection can stimulate release, as reflected in an increased baseline value or an abnormally high interleukin response. Therefore, the PBMC/IL-6 test may only be used if samples have first been shown not to cause

interference. There is no indication that the blood group of the human donors influences the results of the assay.

9.2 Discussion

Comment on and compare the conclusions published in independent peer-reviewed reports or other independent scientific reviews of the proposed test method. The conclusions of such scientific reports and reviews should be compared to the conclusions reached in this submission. Any ongoing evaluations of the proposed test method should be described.

The validation described in this BRD is the first time that such an extensive study for specificity and accuracy using actual medicinal products spiked with endotoxine is carried out. Hence, there are no comparing reports in independent peer-reviewed journals available.

9.3 Results of similar validated method

In cases where the proposed test method is mechanistically and functionally similar to a validated test method with established performance standards, the results of studies conducted with the validated test method subsequent to the ICCVAM evaluation should be included and any impact on the reliability and accuracy of the proposed test method should be discussed.

As mentioned, *in vitro* monocyte activation test methods for the detection of pyrogenic contaminants are being developed over the course of the past two decades. A number of variants have been described, although the underlying principle of each variant remains the same. The test preparation is cultured with monocytoid cells, either as peripheral blood mononuclear cells, PBMCs, (diluted) whole blood or cells of a monocytoid cell line such as MM6. Accuracy and specificity of these test methods are comparable, but in general methods using whole blood, PBMCs and the MM6 cell line appear to perform best (Hoffmann et al, 2005b). Table 9.3.1 summarises the performance of *in vitro* methods presented in the five BRDs and Table 9.3.2 compares the *in vivo* and *in vitro* pyrogen tests regarding their strengths, weaknesses, costs, time, limitations.

However, most studies (as this one) are done with model pyrogens and as yet little experience is available in the field, e.g. as part of the final batch release test-package. Experience and thus confidence in these methods will grow once regulatory authorities approve these methods and more manufacturers start to employ them. Then, on a case by case situation, it should be determined which method is best suited for the actual situation and demonstrates to pick out the appropriate, i.e. pyrogenic batches of the medicinal product.

Table 9.3.1: Summary of the performance of in vitro pyrogen tests based on monocytoid cells (see Tables 7.2.2; 7.2.4; 6.1.3)

Test	System	Read-out	Intralaboratory reproducibility (%)	Interlaboratory reproducibility (%)	Sensitivity (%)	Specificity (%)
WB/IL-6	whole blood	IL-6	DL: 83.3 NL1: 94.4 NL2: 100	DL-NL1: 85.4 DL-NL2: 85.4 NL1-NL2: 92.0	88.9	96.6
WB/IL-1	whole blood	IL-1β	DL: 88. 9 NL1: 95.8 NL2: 94.4	DL-NL1: 72.9 DL-NL2: 81.6 NL1-NL2: 70.2	72.7	93.2
96-wells WB/IL-1 [1]	whole blood	IL-1β	-	DL-NL1: 88.1 DL-NL2: 89.7 NL1-NL2: 91.5	98.8	83.6
CRYO WB/Il-1	cryo whole blood	IL-1β	-	DL-NL1: 91.7 DL-NL2: 91.7 NL1-NL2: 91.7	97.4	81.4
KN CRYO WB/Il-1 [2]	cryo whole blood	IL-1β	-	DL-NL1: 83.3 DL-NL2: 100 NL1-NL2: 83.3	88.9	94.4
PBMC/IL6	PBMC	IL-6	DL: 94.4 NL1: 100 NL2: 94.4	DL-NL1: 84.0 DL-NL2: 86.0 NL1-NL2: 90.0	92.2	95.0
PBMC-CRYO/IL-6 [3]	PBMC	IL-6	-	DL-NL1: 96 DL-NL2: 76 NL1-NL2: 80	93.3	76.7
MM6/IL-6	MM6	IL-6	DL: 100 NL1: 94.4 NL2: 94.4	DL-NL1: 90.0 DL-NL2: 89.6 NL1-NL2: 83.3	95.5	89.8

DL = developing laboratory; NL1, NL2 = naive laboratory 1 and 2
1 = data provided in Section 13 of WB/IL-1 BRD
2 = data provided in Section 13 of CRYO WB/IL-1 BRD
3 = data provided in Section 13 of PBMC/IL-6 BRD

Table amended from Hoffmann et al 2005b; results with THP cells not included

Table 9.3.2: Comparison of the in vivo and in vitro pyrogen tests regarding their strengths, weaknesses, costs, time, limitations

	Rabbit pyrogen test	**BET / LAL**	**In vitro pyrogen test**
Test materials	Liquids	Clear liquids	Liquids, potentially cell preparations, solid materials
Pyrogens covered	All (possible species differences to humans for non-endotoxin pyrogens)	Endotoxin from Gram-negative bacteria	(probably) all
Limit of detection (LPS)	0,5 EU	0,1 EU (some variants down to 0,01 EU)	0,5 EU (validated PM), some variants down to 0,001 EU
Ethical concerns	Animal experiment	About 10% lethality to bled horseshoe crabs	Some assays: blood donation
Costs*	High (200-600$/sample)	Low (50-150$/sample)	Medium (100-350$/sample)
Time required	27 h	45 min	24-30h**
Materials not testable	Short-lived radiochemicals, anesthetics, sedatives, analgetics, chemotherapeutics, immunomodulators, cytokines, corticosteroids	Most biologicals, glucan-containing preparations (herbal medicinal products, cellulose-filtered products), lipids, microsomes, cellular therapeutics	Not known (some of the materials not testable in rabbits require adaptations)
Others	No positive or negative control included, strain differences, stress affects body temperature	Potency of LPS from different bacterial species in mammals not reflected, false-positive for glucans	Possible donor differences, need to exclude hepatitis/HIV and acute infections / allergies of donors, dedifferentiation of cell lines

* = We consulted the laboratories participating in the validation study and a consultant regarding the costs of the tests. The figures we received vary significantly depending on the facility (e.g. industry, contract laboratory, control authority), frequency of testing, specific test requirements, country, etc.

** = interference testing might increase duration by 24 hours

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

10.1 Diminish animal use

Describe how the proposed test method will refine (reduce or eliminate pain or distress), reduce, or replace animal use compared to the reference test method.

Depending on the medicinal product, one of two animal-based pyrogen tests is currently prescribed by the respective Pharmacopoeias, i.e. the rabbit pyrogen test and the bacterial endotoxin test (BET). The rabbit pyrogen test detects various pyrogens but alone the fact that large numbers of animals are required to identify a few batches of pyrogen-containing samples argues against its use. In the past two decades, the declared intention to refine, reduce and replace animal testing, has lowered rabbit pyrogen testing by 80% by allowing to use the BET as an alternative pyrogen test for certain medicinal products. Bacterial endotoxin is the pyrogen of major concern to the pharmaceutical industry due to its ubiquitous sources, its stability and its high pyrogenicity. With the BET, endotoxin is detected by its capacity to coagulate the amoebocyte lysate from the haemolymph of the American horseshoe crab, *Limulus polyphemus*, a principle recognised some 40 years ago (Levin and Bang, 1964). In the US, *Limulus* crabs are generally released into nature after drawing about 20% of their blood and therefore most of these animals survive. However, the procedure still causes mortality of about 30,000 horseshoe crabs per year, which adds to more efficient threats of the horseshoe crab population like its use as bait for fisheries, habitat loss and pollution.

The proposed test method is an alternative for the rabbit test and the BET. By replacing the rabbit test or the BET, the lives of rabbits and horseshoe crabs are spared.

10.2 Continuation of animal use

If the proposed test method requires the use of animals, the following items should be addressed:

10.2.1 *Describe the rationale for the need to use animals and describe why the information provided by the proposed test method requires the use of animals (i.e., cannot be obtained using non-animal methods).*

Not applicable.

10.2.2 *Include a description of the sources used to determine the availability of alternative test methods that might further refine, reduce, or replace animal use for this testing. This should, at a minimum, include the databases searched, the search strategy used, the search date(s), a discussion of the results of the search, and the rationale for not incorporating available alternative methods.*

Not applicable.

10.2.3 *Describe the basis for determining that the number of animals used is appropriate.*

Not applicable.

10.2.4 *If the proposed test method involves potential animal pain and distress, discuss the methods and approaches that have been incorporated to minimize and, whenever possible, eliminate the occurrence of such pain and distress.*
Not applicable.

11 Practical Considerations

11.1 Transferability

Discuss the following aspects of proposed test method transferability. Include an explanation of how this compares to the transferability of the in vivo reference test method and, if applicable, to a comparable validated test method with established performance standards.

In general, the proposed test method is not unlike other bioassays and immunoassays that are performed routinely in many laboratories.

11.1.1 *Discuss the facilities and major fixed equipment needed to conduct a study using the proposed test method.*

No extraordinary facilities are required. General laboratory equipment and analytical instruments for performing immunoassays, e.g. microtiter plate reader and –washer, are sufficient to perform the proposed test method.

11.1.2 *Discuss the general availability of other necessary equipment and supplies.*

All supplies and reagents are readily available on the market. In contrast, availability of sufficient rabbits of adequate weight and in good health for the *in vivo* reference test is sometimes reported a limitation.

It is stressed throughout these studies using PBMC, that only healthy donors not taking any medication must be used for testing pharmaceuticals. Various drugs such as cortisone suppress interleukin release and would therefore exclude a blood donor from the experiment. An infection can stimulate release, as reflected in an increased baseline value or an abnormally high interleukin response.

11.2 Training

Discuss the following aspects of proposed test method training. Include an explanation of how this compares to the level of training required to conduct the in vivo reference test method and, if applicable, a comparable validated test method with established performance standards.

11.2.1 *Discuss the required level of training and expertise needed for personnel to conduct the proposed test method.*

The proposed test method requires personnel trained for general laboratory activities in cell biology and immunochemistry or biochemistry. Techniques they should master are not unlike cell culture (aseptic operations) and immunological techniques (especially ELISA). Such expertise is available in most if not all QC-laboratories.

11.2.2 *Indicate any training requirements needed for personnel to demonstrate proficiency and describe any laboratory proficiency criteria that should be met.*

Personnel should demonstrate that they master the execution of the test. The candidate should demonstrate to meet all the appropriate assay acceptance criteria and yield accurate results (outcome) using selected test items.

11.3 Cost Considerations

Discuss the cost involved in conducting a study with the proposed test method. Discuss how this compares to the cost of the in vivo reference test method and, if applicable, with that of a comparable validated test method with established performance standards.

Two factors contribute to the cost of the proposed test method: cost of the reagents and especially personnel.

Since the proposed test method is relatively more labour-intensive, it is estimated that the cost of the proposed test method is more then the BET or the *in vivo* reference test using rabbits. Obviously, a higher throughput of tests (runs/year) such as in a QC-laboratory of a multi-product facility or in a Contract Research Organization will significantly reduce the costs per assay.

However, especially with pharmaceuticals of biological origin, the proposed test method may be cost-effective, since these products all to often are incompatible with the BET and by their nature preclude the re-use of the rabbits.

11.4 Time Considerations

Indicate the amount of time needed to conduct a study using the proposed test method and discuss how this compares with the in vivo reference test method and, if applicable, with that of a comparable validated test method with established performance standards.

Essentially the test stretches two working days. On day one the testing materials are prepared and incubated overnight with the monocytoid cells. On the second day the amount of excreted cytokines is determined by immunoassay. The total time from start to result is approximately 24 hours.

It is thus concluded that the proposed test method will take more time when compared to the reference tests, either the rabbit test or the BET. It should be noted that rabbits are tested prior to their first use by a sham test.

12 References

List all publications referenced in the submission.

References in bold are included as hardcopies in Appendix B

Barnett, V., Lewis, T. (1984). Outliers in statistical data. 2nd edition, Chichester, US: John Wiley & Sons, pp 171-172.

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

Bleeker, W.K., de Groot, E.M., den Boer, P.J. et al (1994). Measurement of interleukin–6 production by monocytes for in vitro safety testing of hemoglobin solutions. Artif Cells Blood Substit Immobil Biotechnol 22: 835-840.

Clopper, C. J., Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26: 404-413

Co Tui & Schrift, M.H. (1942). A tentative test for pyrogen in infusion fluids. Proc. Soc. Exp. Biol. Med. 49: 320-330.

Cooper JF, Levin J, Wagner Jr. HN (1971). Quantitative comparison of in vitro and *in vivo* methods for the detection of endotoxin. J Lab Clin Med 78: 138-145.

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

Dinarello CA (1999). Cytokines as endogenous pyrogens. J Infect Dis 179 (Suppl 2): S294-304.

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

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

Hansen, E.W. and Christensen, J.D. (1990) Comparison of cultured human mononuclear cells, Limulus amebocyte lysate and rabbits in the detection of pyrogens. J Clin Pharm Ther. 15: 425–433.

Hartung T, Aaberge I, Berthold S et al (2001). Novel pyrogen tests based on the human fever reaction. Altern Lab Anim 29: 99-123.

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

Hoffmann, S., Luederitz-Puechel, U., Montag-Lessing, T., Hartung, T. (2005a). Optimisation of pyrogen testing in parenterals according to different pharmacopoeias by probabilistic modeling. Journal of Endotoxin Research 11(1): 26-31

Hoffmann S, Pertbauer A, Schindler S, et al. (2005b) Validation of novel pyrogen tests based on human monocytoid cell lines. J Immunol Methods, 298(1-2):161-173.

Hothorn, L. A. (1995). Biostatistical analysis of the 'control vs. k treatments' design including a positive control group. In: Testing principls in clinical and preclinical trials, pp 19-26, Stuttgart, Germany: Gustav Fischer Verlag.

Levin J, Bang FB (1964). A description of cellular coagulation in the Limulus. Bull Johns Hopkins Hosp 155: 337-345.

Nakagawa, Y., Maeda, H. and Murai, T. (2002). Evaluation of the *in vitro* pyrogen test system based on proinflammatory cytokine release from human monocytes: comparison with a human whole blood culture test system and with the rabbit pyrogen test. Clin. Diagn. Lab. Immunol. 9: 588-592

Pasare, C. and Medzhitov, R. (2003). Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells. Science 299: 1033-1036.

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

Poole S, Selkirk S, Rafferty B, Meager A, Thorpe R, Gearing A (1989). Assay of pyrogenic contamination in pharmaceuticals by cytokine release. Proceedings of the European Workshop on detection and quantification of pyrogen. Pharmeuropa special Vol 1, november 1989: 17-18.

Poole, S. and Gaines Das, R.E. (2001). Towards a 'human pyrogen test'. J. Parenter. Sci. 6: 63-66.

Poole S, Mistry Y, Ball C et al (2003). A rapid 'one-plate' *in vitro* test for pyrogens. J Immunol Methods 274: 209-220.

Reddi, K., Nair, S.P. et al (1996). Surface-associated material from the bacterium Actinobacillus actinomycetemcomitans contains a peptide which, in contrast to lipopolysaccharide, directly stimulates fibroblast interleukin-6 gene transcription. Eur. J. Biochem. 236: 871-876.

Snedecor, G. W., Cochran, W. G. (1989). Statistical methods, 8th edition, Ames, US: The Iowa State University Press.

Taktak YS, Selkirk S, Bristow AF et al (1991). Assay of pyrogens by interleukin-6 release from monocytic cell lines. J Pharm Pharmacol 43: 578-582.

Tsuchiya S, Yamabe M, Yamaguchi Y et al (1980). Establishment and characterisation of a human acute monocytic leukemia cell line (THP-1). Int J Cancer 26 : 171-176.

Ziegler-Heitbrock HWL, Thiel E, Futterer A et al (1988). Establishment of a human cell line (MONO MAC 6) with characteristics of mature monocytes. Int J Cancer 41: 456-461.

13 CATCH -UP VALIDATION: Human PBMC/IL-6 in vitro Pyrogen Test with Cryopreserved Monocytoid Cells.

13.1 Rationale

The use of cryopreserved cells was investigated separately during a catch-up validation that closely followed the original trial plan (both trial plans are given in Appendix A; see also BRD CRYO WB/IL-1). Cryopreserved PBMC would allow the controlling facility to store (surplus) monocytoid cells from (individual) donors for later use, thereby reducing their waste and increasing flexibility. In addition, it allows for the collection of a larger pool of cells, increasing homogeneity of the reagent. But most of all, it would allow the blood of individual donors to be screened for the absence of adventitious agents before it is actually employed on the laboratory, boosting the safety of the test method.

13.2 Test Method Protocol Components

The method follows the original method protocol (see Appendix A), with the obvious exemption of the cryopreservation of the freshly isolated PBMCs in the presence of 10% (v/v) DMSO. In addition, a 3-donor approach was used; as compared to the original method that required PMBCs from individual 4-donors. A sample is classified as pyrogenic when at least 2 donors graded the sample as pyrogenic. Details are given in the appropriate PBMC-CRYO/IL-6 protocol (Appendix A, Detailed Protocol PBMC-CRYO/IL-6 *In vitro Pyrogen Test Using Freshly Taken or Cryopreserved PBMC (SP+PB var. Novartis; electronic file name: SOP PBMC CRYO IL-6*).

13.3 Substances Used for Validation

The same 10 parenteral drugs used to determine sensitivity and specificity (see table 3.3.1.) were used for the catch-up validation. Again, each test item was tested after spiking at its individual MVD, thus came with their own specific set of 5 endotoxin spike solutions: 0.0, 0.25, 0.5, 0.5 and 1.0 EU/ml. The test items were assessed with 5 different endotoxin levels at 3 independent test facilities, yielding a total of 150 data points, biometrically considered to be sufficient for further analysis. As all the test results qualified according to quality criteria, the maximum number was analysed.

13.4 Test Method Accuracy

To assess **accuracy** of the proposed test method 10 substances (listed in table 3.3.1), were spiked with five different concentrations of the WHO-LPS (one of which is negative). To permit the application of the chosen PM, each drug was diluted to its individual MVD, which has been calculated using the ELC to that drug (listed in table 3.3.1.) Each substance had their own specific set of endotoxin spike solution, ensuring that after spiking the undiluted drug, it contained 0.0; 0.25; 0.5; 0.5 and 1.0 EU/ml respectively when tested at its MVD. A detailed description of the sample preparation was supplied to the three independent laboratories (as shown in table 5.1.2). To put more weight to the result of this part of the validation, the spikes were blinded and coded by QA ECVAM. Accuracy was assessed by applying the PM to the results and evaluating the concordance in a two by two table.

Table 13.4.1: Results of the validation study of 10 drugs, spiked with WHO-LPS at 0.0, 0.25, 0.5, 0.5 and 1.0 IU/ml, respectively and tested in 3 different laboratories. Samples and spikes were blinded. Classifications after applying the PM (compare to fig. 5.2.7).

drug (code)	**spike EU/ml**	**"truth"**	**results** Novartis (Basel, Ch)	NIBSC (UK)	PEI (Ger)
Beloc (BE)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	**0**	**0**	1
	0.50	1	0	0	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Binotal (BI)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	1	1	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Ethanol 13% (ET)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	1	1	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Fenistil (FE)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	**0**	**0**	1
	0.50	1	0	0	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Glucose 5% (GL)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	**0**	**0**	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
"Drug A" 0.9% NaCl (LO)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	**0**	**0**	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
MCP (ME)	**0.00**	**0**	**0**	**0**	**0**
	0.25	**0**	**0**	**0**	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1

drug (code)	spike EU/ml	"truth"	results Novartis (Basel, Ch)	NIBSC (UK)	PEI (Ger)
"Drug B" 0.9% NaCl (MO)	0.00	0	0	0	0
	0.25	0	0	0	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Orasthin (OR)	0.00	0	0	0	0
	0.25	0	0	0	1
	0.50	1	1	1	1
	0.50	1	1	1	1
	1.00	1	1	1	1
Sostril (SO)	0.00	0	0	0	0
	0.25	0	0	0	1
	0.50	1	0	1	1
	0.50	1	0	1	1
	1.00	1	1	1	1

"0"denotes "non-pyrogenic"; "1" denotes "pyrogenic".

Table 13.4.2: 2x2 contingency table. PM applied to the PBMC-CRYO/IL-6 test result of 10 different substances assessed in three different laboratories. Results of each laboratory separately (DL, NL1 and NL2= Novartis, NIBSC and PEI respectively).

Results DL		True status of samples +	-	Total
PM	+	26	2	28
	-	4	18	22
Total		30	20	**50**

Results NL1		True status of samples +	-	Total
PM	+	28	2	30
	-	2	18	20
Total		30	20	**50**

Results NL2		True status of samples +	-	Total
PM	+	30	10	40
	-	0	10	10
Total		30	20	**50**

Table 13.4.2: 2x2 contingency table. PM applied to the PBMC-CRYO/IL-6 test result of 10 different substances assessed in three different laboratories (from table 13.4.1).

	True status of samples +	-	**Total**
PM +	84	14	**98**
-	6	46	**52**
Total	**90**	**60**	**150**

The overall specificity of the PBMC-CRYO/IL-6 assay is 76.7% (46/(46+14)*100). The overall sensitivity equals 93.3% (84/(84+6) *100%). Within the laboratories, specificity varied from 50% in one participating laboratory up to 90% in the other two laboratories. The sensitivity varied from 86.7% up to 93.3% and 100%.

Table 13.4.3: Specificity and sensitivity of the PBMC/IL-6 assay as determined from table 13.4.2.

	N total	**N correctly identified**	**proportion**	**95% CI lower limit**	**95% CI upper limit**
Specificity (Sp)	60	46	76.7%	64.7%	87.5%
Sensitivity (Se)	90	84	93.3%	85.6%	97.4%

13.5 Test Method Reliability (Reproducibility)

The interlaboratory reproducibility of the PBMC-CRYO/IL-6 method was assessed from the results of the catch-up validation testing 10 substances spiked with 5 separate spikes. The reproducibility varied from 76% to 80% and 96% between two laboratories. All three laboratories found the same result for 38 out of 50 samples.

Table 13.5.1: Interlaboratory reproducibility: Assessed by testing of 10 substances, spiked 5 times. One run of 50 samples by three different laboratories.

Laboratories	**Interlaboratory Reproducibility**	**Number of equal predictions**
DL - NL1	96%	48 / 50
DL - NL2	76%	38 / 50
NL1 – NL2	80%	40 / 50
Mean	84%	
same result in all labs	76%	38 / 50

DL =Novartis; NL1 = NIBSC; NL2 = PEI

13.6 Conclusion

It appears that accuracy and reproducibility of the test method do not decrease when cryopreserved cells replace fresh PBMCs. The sensitivity, which is well above 90%, is considered quite satisfactory for both methods. The somewhat lower results for specificity can fully explained by several false positives in just one of the laboratories. In the other two laboratories the specificity is still 90%, with a corresponding reproducibility of 96%.

14 Supporting Materials (Appendices)

14.1 Standard operating procedure (SOP) of the proposed method

Provide the complete, detailed protocol for the proposed test method.
Appendix A includes the PBMC/IL-6 test method protocol using fresh blood cells as used throughout the studies described in section 5 of the current BRD.
In addition, Appendix A includes the PBMC-CRYO/IL-6 test method protocol using cryopreserved blood cells. The protocol was used in catch-up validation study described in Section 13 of this BRD.
The trial plans of both studies are also included in Appendix A.

14.2 Standard operating Procedure (SOP) of the reference method

Provide the detailed protocol(s) used to generate reference data for this submission and any protocols used to generate validation data that differ from the proposed protocol.

14.3 Publications

Provide copies of all relevant publications, including those containing data from the proposed test method, the in vivo reference test method, and if applicable, a comparable validated test method with established performance standards.

Not all of the publications cited in the BRD (see section 12) are included as hardcopies in Appendix B, e.g. publications on statistical methods are not given.

Remark: The same set of hardcopies was included into Appendix B of all of the 5 submitted BRDs. Therefore some of the publications in Appendix B might not be referenced in the current BRD nor included in the list of the publications in section 12. Three general publications were added, which are not cited in any BRD but might be useful as background information to the validation study: the ECVAM publications on validation (Balls et al, 1995; Curren et al, 1995; Hartung et al, 2004). Several publications were included, which either give more background information on the human fever reaction or report on specific studies using *in vitro* pyrogen tests.

List of hard copies

Andrade SS, Silveira RL, Schmidt CA, Junior LB, Dalmora SL. (2003) Comparative evaluation of the human whole blood and human peripheral blood monocyte tests for pyrogens. Int J Pharm. Oct 20;265(1-2):115-24.

Balls et al. (1995) Practical aspects of the validation of toxicity test procedures. ECVAM Workshop Report 5. ATLA 23, 129-147.

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

Bleeker, W.K., de Groot, E.M., den Boer, P.J. et al (1994). Measurement of interleukin–6 production by monocytes for in vitro safety testing of hemoglobin solutions. Artif Cells Blood Substit Immobil Biotechnol 22: 835-840.

Carlin G, Viitanen E. (2003) In vitro pyrogenicity of a multivalent vaccine: Infanrix. PHARMEUROPA, 15(3), 418-423.

Curren et al (1995) The Role of prevalidation in the development, validation and acceptance of Alternative Methods. ECVAM Prevalidation Task Force Report 1. ATLA 23, 211-217.

De Groote, D., Zangerle, P.F., Gevaert, Y., Fassotte, M.F., Beguin, Y., Noizat-Pirenne, F., Pirenne, J., Gathy, R., Lopez, M., Dehart I., (1992). Direct stimulation of cytokines (IL-1 beta, TNF-alpha, IL-6, IL-2, IFN-gamma and GM-CSF) in whole blood. I. Comparison with isolated PBMC stimulation. Cytokine 4, 239.

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

Dinarello CA (1999). Cytokines as endogenous pyrogens. J Infect Dis 179 (Suppl 2): S294-304.

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

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

Fennrich S, Fischer M, Hartung T, Lexa P, Montag-Lessing T, Sonntag H-G, Weigandt M und Wendel A. (1999). Detection of endotoxins and other pyrogens using human whole blood. Dev. Biol. Standards 101:131-139

Gaines Das et al (2004). Monocyte activation test for pro-inflammatory and pyrogenic contaminants of parenteral drugs: test design and data analysis. J. of Immunological Methods 288, 165-177.

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

Hansen, E.W. and Christensen, J.D. (1990) Comparison of cultured human mononuclear cells, Limulus amebocyte lysate and rabbits in the detection of pyrogens. J Clin Pharm Ther. 15: 425–433.

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

Hartung T, Aaberge I, Berthold S et al (2001). Novel pyrogen tests based on the human fever reaction. Altern Lab Anim 29: 99-123.

Hartung et al (2004) A modular approach to the ECVAM principles on test validity. ATLA 32, 467-472.

Hoffmann, S., Luederitz-Puechel, U., Montag-Lessing, T., Hartung, T. (2005). Optimisation of pyrogen testing in parenterals according to different pharmacopoeias by probabilistic modeling. Journal of Endotoxin Research 11(1): 26-31

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

Jahnke M, Weigand, Sonntag H-G. (2000). Comparative testing for pyrogens in parenteral drugs using the human whole blood pyrogen test, the rabbit in vivo pyrogen test and the LAL test. European Journal of Parenteral Science 5(2):39-44

Nakagawa Y, Maeda H, Murai T. (2002) Evaluation of the in vitro pyrogen test system based on proinflammatory cytokine release from human monocytes: comparison with a human whole blood culture test system and with the rabbit pyrogen test. Clin Diagn Lab Immunol. May;9(3):588-97.

Pasare, C. and Medzhitov, R. (2003). Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells. Science 299: 1033-1036.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. (1998) The detection of pyrogens in blood products using an ex vivo whole blood culture assay. J Immunoassay. May-Aug;19(2-3):95-111.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. (1999) Differentiation between endotoxin and non-endotoxin pyrogens in human albumin solutions using an ex vivo whole blood culture assay. J Immunoassay. Feb-May; 20(1-2):79-89.

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

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

Poole S, Selkirk S, Rafferty B, Meager A, Thorpe R, Gearing A (1989). Assay of pyrogenic contamination in pharmaceuticals by cytokine release. Proceedings of the European Workshop on detection and quantification of pyrogen. Pharmeuropa special Vol 1, November 1989: 17-18.

Poole S, Musset MV (1989). The International Standard for Endotoxin: evaluation in an international collaborative study. J Biol Stand 17: 161-171.

Poole S, Mistry Y, Ball C et al (2003). A rapid 'one-plate' in vitro test for pyrogens. J Immunol Methods 274: 209-220.

Ray A, Redhead K, Selkirk S, Polle S (1991) Variability in LPS composition, antigenicity and reactogenicity of phase variants of Bordetella pertussis. FEMS Microbiology Letters 79, 211-218.

Reddi, K., Nair, S.P. et al (1996). Surface-associated material from the bacterium Actinobacillus actinomycetemcomitans contains a peptide which, in contrast to lipopolysaccharide, directly stimulates fibroblast interleukin-6 gene transcription. Eur. J. Biochem. 236: 871-876.

Schindler S, Bristow A, Cartmell T, Hartung T and Fennrich S. (2003). Comparison of the reactivity of human and rabbit blood towards pyrogenic stimuli. ALTEX 20:59-63.

Schindler et al (2004) Cryopreservation of human whole blood for pyrogenicity testing. Journal of Immunological Methods 294, 89–100

Schins RP, van Hartingsveldt B, Borm PJ. Ex vivo cytokine release from whole blood. A routine method for health effect screening. Exp Toxicol Pathol. 1996 Nov; 48(6):494-6.

Spreitzer I, Fischer M, hartzsch K, Lüderitz-Püschel U and Montag T. (2000). Comparative Study of Rabbit Pyrogen Test and Human whole Blood Assay on Human Serum Albumin. ALTEX 19 (Suppl. 1):73-75

Taktak YS, Selkirk S, Bristow AF et al (1991). Assay of pyrogens by interleukin-6 release from monocytic cell lines. J Pharm Pharmacol 43: 578-582.

Tsuchiya S, Yamabe M, Yamaguchi Y et al (1980). Establishment and characterisation of a human acute monocytic leukemia cell line (THP-1). Int J Cancer 26 : 171-176.

Ziegler-Heitbrock HWL, Thiel E, Futterer A et al (1988). Establishment of a human cell line (MONO MAC 6) with characteristics of mature monocytes. Int J Cancer 41: 456-461.

14.4 Original data

Include all available non-transformed original data for both the proposed test method, the in vivo reference test method, and if applicable, a comparable validated test method with established performance standards.

NOTE: The original data of the ELISA-plate reader were collected by S.Hoffman and ECVAM. These are available on the CD which goes with the BRD.

14.5 Performance standards

If appropriate performance standards for the proposed test method do not exist, performance standards for consideration by NICEATM and ICCVAM may be proposed. Examples of established performance standards can be located on the ICCVAM / NICEATM web site at http://iccvam.niehs.nih.gov.

BRD: PBMC/IL-6 March, 2006

APPENDIX A

Trial plan "Comparison And Validation Of Novel Pyrogen Tests Based On The Human Fever Reaction" Acronym: Human (e) Pyrogen Test

BRD: PBMC/IL-6 March, 2006

Detailed protocol PBMC/IL-6: "In Vitro Pyrogen Test Using PBMC (SP+PB var. Novartis) 03 10 02

Trial plan "Catch-up Validation of Novel Pyrogen Tests Based on the Human Fever Reaction"

Detailed protocol PBMC/IL-6 CRYO: In vitro Pyrogen Test Using Freshly Taken or Cryopreserved PBMC (SP+PB var. Novartis)

APPENDIX B

Not all of the publications cited in the BRD (see section 12) are included as hardcopies in Appendix B, e.g. publications on statistical methods are not given.

Remark: The same set of hardcopies was included into Appendix B of all of the 5 submitted BRDs. Therefore some of the publications in Appendix B might not be referenced in the current BRD nor included in the list of the publications in section 12. Three general publications were added, which are not cited in any BRD but might be useful as background information to the validation study: the ECVAM publications on validation (Balls et al, 1995; Curren et al, 1995; Hartung et al, 2004). Several publications were included, which either give more background information on the human fever reaction or report on specific studies using *in vitro* pyrogen tests.

List of hard copies

Andrade SS, Silveira RL, Schmidt CA, Junior LB, Dalmora SL. (2003) Comparative evaluation of the human whole blood and human peripheral blood monocyte tests for pyrogens. Int J Pharm. Oct 20;265(1-2):115-24.

Balls et al. (1995) Practical aspects of the validation of toxicity test procedures. ECVAM Workshop Report 5. ATLA 23, 129-147.

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

Bleeker, W.K., de Groot, E.M., den Boer, P.J. et al (1994). Measurement of interleukin–6 production by monocytes for in vitro safety testing of hemoglobin solutions. Artif Cells Blood Substit Immobil Biotechnol 22: 835-840.

Carlin G, Viitanen E. (2003) In vitro pyrogenicity of a multivalent vaccine: Infanrix. PHARMEUROPA, 15(3), 418-423.

Curren et al (1995) The Role of prevalidation in the development, validation and acceptance of Alternative Methods. ECVAM Prevalidation Task Force Report 1. ATLA 23, 211-217.

De Groote, D., Zangerle, P.F., Gevaert, Y., Fassotte, M.F., Beguin, Y., Noizat-Pirenne, F., Pirenne, J., Gathy, R., Lopez, M., Dehart I., (1992). Direct stimulation of cytokines (IL-1 beta, TNF-alpha, IL-6, IL-2, IFN-gamma and GM-CSF) in whole blood. I. Comparison with isolated PBMC stimulation. Cytokine 4, 239.

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

Dinarello CA (1999). Cytokines as endogenous pyrogens. J Infect Dis 179 (Suppl 2): S294-304.

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

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

Fennrich S, Fischer M, Hartung T, Lexa P, Montag-Lessing T, Sonntag H-G, Weigandt M und Wendel A. (1999). Detection of endotoxins and other pyrogens using human whole blood. Dev. Biol. Standards 101:131-139

Gaines Das et al (2004). Monocyte activation test for pro-inflammatory and pyrogenic contaminants of parenteral drugs: test design and data analysis. J. of Immunological Methods 288, 165-177.

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

Hansen, E.W. and Christensen, J.D. (1990) Comparison of cultured human mononuclear cells, Limulus amebocyte lysate and rabbits in the detection of pyrogens. J Clin Pharm Ther. 15: 425–433.

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

Hartung T, Aaberge I, Berthold S et al (2001). Novel pyrogen tests based on the human fever reaction. Altern Lab Anim 29: 99-123.

Hartung et al (2004) A modular approach to the ECVAM principles on test validity. ATLA 32, 467-472.

Hoffmann, S., Luederitz-Puechel, U., Montag-Lessing, T., Hartung, T. (2005). Optimisation of pyrogen testing in parenterals according to different pharmacopoeias by probabilistic modeling. Journal of Endotoxin Research 11(1): 26-31

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

Jahnke M, Weigand, Sonntag H-G. (2000). Comparative testing for pyrogens in parenteral drugs using the human whole blood pyrogen test, the rabbit in vivo pyrogen test and the LAL test. European Journal of Parenteral Science 5(2):39-44

Nakagawa Y, Maeda H, Murai T. (2002) Evaluation of the in vitro pyrogen test system based on proinflammatory cytokine release from human monocytes: comparison with a human whole blood culture test system and with the rabbit pyrogen test. Clin Diagn Lab Immunol. May;9(3):588-97.

Pasare, C. and Medzhitov, R. (2003). Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells. Science 299: 1033-1036.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. (1998) The detection of pyrogens in blood products using an ex vivo whole blood culture assay. J Immunoassay. May-Aug;19(2-3):95-111.

Pool EJ, Johaar G, James S, Petersen I, Bouic P. (1999) Differentiation between endotoxin and non-endotoxin pyrogens in human albumin solutions using an ex vivo whole blood culture assay. J Immunoassay. Feb-May; 20(1-2):79-89.

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

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

Poole S, Selkirk S, Rafferty B, Meager A, Thorpe R, Gearing A (1989). Assay of pyrogenic contamination in pharmaceuticals by cytokine release. Proceedings of

the European Workshop on detection and quantification of pyrogen. Pharmeuropa special Vol 1, November 1989: 17-18.

Poole S, Musset MV (1989). The International Standard for Endotoxin: evaluation in an international collaborative study. J Biol Stand 17: 161-171.

Poole S, Mistry Y, Ball C et al (2003). A rapid 'one-plate' in vitro test for pyrogens. J Immunol Methods 274: 209-220.

Ray A, Redhead K, Selkirk S, Polle S (1991) Variability in LPS composition, antigenicity and reactogenicity of phase variants of Bordetella pertussis. FEMS Microbiology Letters 79, 211-218.

Reddi, K., Nair, S.P. et al (1996). Surface-associated material from the bacterium Actinobacillus actinomycetemcomitans contains a peptide which, in contrast to lipopolysaccharide, directly stimulates fibroblast interleukin-6 gene transcription. Eur. J. Biochem. 236: 871-876.

Schindler S, Bristow A, Cartmell T, Hartung T and Fennrich S. (2003). Comparison of the reactivity of human and rabbit blood towards pyrogenic stimuli. ALTEX 20:59-63.

Schindler et al (2004) Cryopreservation of human whole blood for pyrogenicity testing. Journal of Immunological Methods 294, 89–100

Schins RP, van Hartingsveldt B, Borm PJ. Ex vivo cytokine release from whole blood. A routine method for health effect screening. Exp Toxicol Pathol. 1996 Nov; 48(6):494-6.

Spreitzer I, Fischer M, hartzsch K, Lüderitz-Püschel U and Montag T. (2000). Comparative Study of Rabbit Pyrogen Test and Human whole Blood Assay on Human Serum Albumin. ALTEX 19 (Suppl. 1):73-75

Taktak YS, Selkirk S, Bristow AF et al (1991). Assay of pyrogens by interleukin-6 release from monocytic cell lines. J Pharm Pharmacol 43: 578-582.

Tsuchiya S, Yamabe M, Yamaguchi Y et al (1980). Establishment and characterisation of a human acute monocytic leukemia cell line (THP-1). Int J Cancer 26 : 171-176.

Ziegler-Heitbrock HWL, Thiel E, Futterer A et al (1988). Establishment of a human cell line (MONO MAC 6) with characteristics of mature monocytes. Int J Cancer 41: 456-461.

APPENDIX C

List of abbreviations and definitions

Accuracy	The ability of a test system to provide a test result close to the accepted reference value for a defined property.
BET	The bacterial endotoxin test is used to detect or quantify endotoxins of gram-negative bacterial origin using amoebycte lysate from horseshoe crab (*Limulus polyphemus* or *Tachypleus tridentatus*
BRD	Background Review Document
CRYO WB/IL-1	Whole blood assay (using cryopreserved blood) with IL-1 as endpoint
CV	coefficient of variation
DL	Developing laboratory = laboratory which developed the method or the most experienced laboratory
ELC	Endotoxin limit concentration; maximum quantity of endotoxin allowed in given parenterals according to European Pharmacopoeia
Endotoxins	Endotoxins are a group of chemically similar cell-wall structures of Gram-negative bacteria, i.e. lipopolysaccharides
ELISA	Enzyme linked immunosorbent assay
EU/ml	European Units per ml
IL-1	interleukin 1
IL-6	interleukin 6
Intralaboratory reproducibility	A determination of the extent that qualified people within the same laboratory can independently and successfully replicate results using a specific protocol at different times.
Interlaboratory reproducibility	A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is also referred to as between-laboratory reproducibility.
KN	University of Konstanz (Konstanz, Germany), developing laboratory WB/IL-1 and CRYO WB/IL-1
LPS	lipopolysaccharides
MM6	MONO MAC-6 cell line

MM6/IL-6	In vitro pyrogen test using MM6 cell line and IL-6 release as an endpoint
MVD	Maximum valid dilution; the MVD is the quotient of the ELC and the detection limit
NIBSC	National Institute for Biological Standards and Control (London, UK), developing laboratory for WB/IL-6
NL	naïve laboratory = laboratory with non or minor experience with the method
NPC	negative product control (clean, released lot of the nominated product under test)
Novartis	Novartis (Basel, Switzerland), developing laboratory PBMC/IL-6
OD	optical density
PBMC	Peripheral blood mononuclear cells
PBMC/IL-6	In vitro pyrogen test using fresh peripheral blood mononuclear cells and IL-6 release as endpoint
PBMC-CRYO/IL-6	In vitro pyrogen test using cryopreserved peripheral blood mononuclear cells and IL-6 release as endpoint
PEI	Paul-Ehrlich Institut (Langen, Germany), participating laboratory
PM	prediction model = is an explicit decision-making rule for converting the results of the in vitro method into a prediction of in vivo hazard
PPC	positive product control (product under test spiked with 0.5 EU/ml of WHO-LPS (code 94/580)
Prevalidation study	A prevalidation study is a small-scale inter-laboratory study, carried out to ensure that the protocol of a test method is sufficiently optimised and standardised for inclusion in a formal validation study. According to the ECVAM principles, the prevalidation study is divided into three phases: protocol refinement, protocol transfer and protocol performance (Curren et al, ATLA 23, 211-217).
Pyrogens	fever-causing materials
Pyrogens, endogenous	endogenous pyrogens are messenger substances released by blood cells reacting to pyrogenic materials; e.g. IL-1, IL-6, TNF-α, prostaglandin E_2
Pyrogens, exogenous	exogenous pyrogens derive from bacteria, viruses, fungi or from the host himself
Reliability	Measures of the extent to which a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is

	assessed by calculating intra- and interlaboratory reproducibility and intra-laboratory repeatability.
Relevance	Relevance of a test method describes whether it is meaningful and useful for a particular purpose. It is the extent to which the measurement result and uncertainty can accurately be interpreted as reflecting or predicting the biological effect of interest.
Repeatibility	Repeatability describes the closeness of agreement between test results obtained within a single laboratory when the procedure is performed independently under repeatability conditions, i.e. in a set of conditions including the same measurement procedure, same operator, same measuring system, same operating conditions and same location, and replicated measurements over a short period of time.
RIVM	National Institute of Public Health and the Environment (Bilthoven, The Netherlands), developing laboratory MM6/IL-6 method
Sensitivity	Sensitivity is the proportion of all positive/active substances that are correctly classified by a test method.
Specificity	Specificity is proportion of all negative/inactive substances that are correctly classified by a test method.
TMB	chromogenic substrate 3,3′,5,5′ -tetramethylbenzidine
TNF-α	tumour necrosis factor-α
USP	US Pharmacopoeia
Validation	Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose
Validation study	A validation study is a large-scale interlaboratory study, designed to assess the reliability and relevance of an optimised method for a particular purpose
WB/IL-1	Whole blood assay (using fresh blood) with IL-1 release as endpoint
WB/IL-6	Whole blood assay (using fresh blood) with IL-6 release as endpoint
WHO	World Health Organization