Appendix D3

SACATM Comments: ICCVAM Evaluation of *In Vitro* Pyrogen Test Methods

[This Page Intentionally Left Blank]

The following is excerpted from the final minutes of the SACATM meeting convened on June 12, 2007. The full meeting minutes will be available online at http://ntp-server.niehs.nih.gov/ntpweb/index.cfm?objectid=AF6CC417-F1F6-975E-75B5F3FF7DF1CDDC.

Overview of the ICCVAM Evaluation of *In Vitro* Pyrogen Test Methods

Dr. Richard McFarland, U.S. Food and Drug Administration (FDA), ICCVAM Pyrogenicity Working Group (PWG) Chair, presented an update on ICCVAM's ongoing evaluation of five *in vitro* human cell-based pyrogen test methods nominated for review by the European Centre for the Validation of Alternative Methods (ECVAM). Pyrogenicity is defined as an increase in body temperature following the release of pro-inflammatory cytokines [e.g., interleukin (IL)-1, IL-6, and tumor necrosis factor-α (TNF-α)] by leukocytes. Pyrogens may be found in processing and packaging materials, chemicals, or parenteral pharmaceuticals, biologicals, and medical devices. Bacterial endotoxin, a component of the outer cell wall of Gram-negative bacteria, is one of the most potent pyrogenic materials. Pyrogen testing is important to prevent the introduction of endotoxin or non-endotoxin pyrogen-contaminated products into humans or animals.

Currently there are two accepted pyrogen tests. The Rabbit Pyrogen Test (RPT), which measures a temperature rise in rabbits injected with a test substance, can detect both endotoxin and non-endotoxin pyrogens. The Bacterial Endotoxin Test (BET), also referred to as the *Limulus* Amoebocyte Lysate (LAL) Test, detects endotoxin by its ability to activate a serine-protease catalytic cascade.

In June 2005, ECVAM submitted background review documents (BRDs) on five methods for consideration by NICEATM as replacements for the RPT. The methods are:

- Human Whole Blood (WB)/Interleukin (IL)-1 *In Vitro* Pyrogen Test
- Human WB/IL-1 *In Vitro* Pyrogen Test: Application of Cryopreserved Human WB
- Human WB/IL-6 *In Vitro* Pyrogen Test
- Human Peripheral Blood Mononuclear Cell (PBMC)/IL-6 *In Vitro* Pyrogen Test
- *In Vitro* Pyrogen Test using the monocytoid cell line, Mono Mac 6 (MM6)/IL-6

Before describing the evaluation process, Dr. McFarland listed the members of the PWG, provided a time line for the various activities connected with the evaluation process, and described the ICCVAM acceptance and validation criteria for alternative test methods.

Following a prescreen evaluation, NICEATM requested additional information and clarification from ECVAM in regard to the data provided in their BRDs. ECVAM submitted revised BRDs that addressed these requests. Subsequently, ICCVAM prepared a draft ICCVAM BRD that contained a comprehensive review of all available data and information regarding the usefulness and limitations of the five alternative *in vitro* pyrogen test methods and described the current validation status of the test methods including their relevance, reliability, scope of substances tested, and the availability of a standardized test method protocol for each test method.

The major difference among the five test methods is the cell types used; the methodology used for the test methods is very similar. Briefly, the test substance is applied to cultures of the specific human-derived cells, which are then incubated for 16-24 hr. The concentration of pro-inflammatory cytokines (e.g., IL-1β, IL-6) is quantified via a cytokine-specific enzyme-linked immunosorbent assay (ELISA). The endotoxin activity of a test substance is calculated by comparing the induced cytokine release with that induced by the endotoxin standard.

The test methods were reviewed for their ability to detect the presence of Gram-negative endotoxin when several parenteral pharmaceuticals were spiked with the endotoxin standard at several different concentrations. The reference pharmaceuticals were considered positive for endotoxin if the endotoxin content was > 0.5 endotoxin units (EU)/mL. Differences were found in the performance of the five test methods. Based on the information contained in the BRD, ICCVAM developed draft recommendations for the use, formulated draft performance standards and draft test method protocols for each test method, and identified proposed future studies.

ICCVAM's draft recommendations on test method uses and limitations was that, based on the validation studies with a limited number of pharmaceuticals, there is sufficient information to substantiate the use of these test methods for the detection of pyrogenicity mediated by Gram-negative endotoxins in materials that are currently tested in the RPT, subject to product-specific validation to demonstrate equivalency. Further, ICCVAM's draft recommendations stated that although the five *in vitro* test methods may be capable of detecting a wider range of pyrogens than was tested, the data in the BRDs do not support this broader application. One limitation of the validation study was the lack of a direct comparison of the results for the same test substances in the proposed *in vitro* test methods versus the RPT.

ICCVAM also provided draft recommendations for performance standards for these five *in vitro* test methods for consideration by the peer review panel and for public comment; the purpose of performance standards are to ensure that any proposed mechanistically and functionally similar proposed test method meets acceptable standards. Performance standards include essential test method components based upon common structural, functional, and procedural elements that should be included in the protocol of a mechanistically and functionally similar proposed test method; recommended reference substances for evaluating the relevance and reliability of the proposed test method and the performance characteristics (relevance and reliability values) that should be met or exceeded. ICCVAM also recommended draft standardized protocols that were based on those used in the ECVAM validation study. Finally, ICCVAM recommended future studies that included the testing of a broader range of pyrogenic materials under conditions where the *in vitro* pyrogen test(s) and the RPT were run in parallel to be able to directly compare the results.

Peer Panel Report

ICCVAM and NICEATM held a peer review panel meeting on February 6, 2007, to review the five *in vitro* pyrogenicity test methods. Dr. Karen Brown, DRL Pharma and

Pair O'Doc's Enterprises, chair of the peer panel, said the task was daunting because the panel was tasked to complete the evaluation of the five *in vitro* test methods in one day. She recognized the hard work and diligence of the panel.

The charge to the peer review panel was to review the draft BRDs for completeness, assess whether each applicable criterion for validation and acceptance of the test method had been appropriately addressed, and consider whether the information in the BRD supported the draft ICCVAM recommendations for the draft standardized protocols, the draft test method performance standards, and the draft proposed future studies.

The panel concluded that the explanation in the BRD of the usefulness and limitations of the *in vitro pyrogen*icity test methods and of the description of the current validation status of these methods was sufficient. However, they identified a number of deficiencies in the BRD, which are briefly described below.

1. There were some sections where additional details would have improved the document. For example, the panel wanted information included about (1) the number of RPTs conducted per year to evaluate bacterial endotoxin, (2) the number of rabbits used for *pyrogen*icity testing per year, and (3) the costs and logistical considerations for either setting up the cell culture for the MM6 test or obtaining human blood for the other tests.

2. The rationale for selecting the test substances for evaluating the five *in vitro* test methods was flawed because it did not represent the range of products tested for bacterial endotoxin using the RPT and seven of the 10 substances were not tested in the RPT but rather in the BET. For example, no biologicals or medical devices were evaluated. The panel felt that the number of substances tested in the validation study was not adequate to evaluate whether a specific test method could replace the RPT.

3. The *in vivo* RPT reference data were limited to one strain of rabbit tested in one laboratory by one protocol using two sources of bacterial endotoxin.

4. The evaluation of the relevance of each test method was adequately demonstrated and discussed in the BRDs, but was limited by the ability to judge a positive versus a negative response based on 0.5 endotoxin units (EU)/ mL. Since samples were only spiked with bacterial endotoxin, the relevance was only demonstrated for the detection of this type of pyrogen, and there was no evaluation for the ability to detect non-endotoxin pyrogens.

5. The discussion on concordance in the RPT is speculative because there was no parallel testing with the RPT, and the RPT performance was modeled statistically.

6. The whole blood IL-1 test is inadequate because there were too many false positives and false negatives; however the IL-6 assay appeared to perform better.

It would have been more appropriate to compare these *in vitro* tests directly with the BET, since only bacterial endotoxin samples were used.

7. Test method reliability was acceptable in both within and between laboratory studies; however, a quantitative assessment of intra- and inter-laboratory variability would have been more informative. A statistical assessment providing acceptability criteria should have been performed to test the hypothesis that there were no differences among groups.

8. The assessment of test method reliability had the following deficiencies:

    a. There was a high exclusion rate for individual runs of the whole blood IL-1 assay due to excessive variability among the four replicates.
    b. The agreement across three validation laboratories was only 57% for the whole blood IL-1 assay.
    c. The same subset of drugs tested for sensitivity and specificity should have been tested for reliability.

Most of the panel agreed that application of the validation criteria to determine the usefulness and limitations of these test methods to replace the RPT under conditions where the test was for the presence of Gram-negative endotoxin was adequately addressed in the BRDs.

The panel concluded that the usefulness of the test methods to detect Gram-negative endotoxin was not assessed properly to determine their concordance with the RPT or to compare their relevance with the BET. The assessment of the usefulness was limited because non-endotoxin pyrogens were not included, and the pure form of the test materials may stimulate cytokine production.

The panel agreed that the BRDs did support the proposed standardized test method protocols if the list of its inadequacies were fully addressed. The panel noted that to reduce variability, similar acceptance criteria must be used for multiple blood donors and similar exclusion rules must be used for each test method. They recommended that a more specific protocol be developed that details recruitment of human blood donors, selection criteria for donors, as well as conditions for veinipuncture.

The panel concluded that the test method performance standards were not supported by the BRD. Statements about the five methods' accuracy and reliability were not supported because two assays demonstrated false-positive results greater than 16 % and the *in vitro* test methods should have been compared to both the BET and RPT. Also, the panel thought that the small list of substances was inadequate to assess whether these test methods could replace the RPT. Test substances need to include all classes of endotoxins as well as non-endotoxin pyrogens.

The panel agreed that additional studies should be performed, and that ICCVAM should consider their comments and recommendations. They suggested (1) establishment of a

repository of clinically identified pyrogens to use in future validation studies, (2) inclusion of both endotoxin and non-endotoxin pyrogens in future validation studies, (3) prospective comparison of any *in vitro* tests with the RPT and BET, and (4) evaluation of IL-1 and IL-6 levels in the *in vitro* tests and their correlation with levels produced in rabbits exposed to similar levels of endotoxin.
Overall, the peer review panel concluded that these five test methods could be applicable for a wider range of pyrogens and test materials if they were adequately validated for such uses. It is important to recognize that, despite the panel's concerns about the performance of these five *in vitro* test methods, the FDA has a formal process for materials regulated under 21CFR610.9 (e.g., parenteral drugs) that allows drug manufacturers to qualify *in vitro test* methods for identifying Gram-negative endotoxin, on a case-by-case basis.

Public comments:
Dr. Freedman identified the written comments submitted by Physicians Committee for Responsible Medicine (PCRM).

Ms. Kristy Stoick, PCRM, said her organization submitted written comments after the peer review panel meeting. PCRM was disappointed with the ICCVAM draft recommendations and the peer review panel report. Since federal regulations specify that these methods must undergo product specific validation for pyrogenicity, she encouraged SACATM to recommend that ICCVAM help facilitate further development of these methods by companies so the regulatory community can begin to use them as soon as possible. She did not support additional *in vivo* validation studies.

Dr. Thomas Hartung, ECVAM, joined the public for this specific agenda item because of a conflict of interest as a patent holder for the methods. Three of the *in vitro* test methods were based on his research and he had coordinated the validation study prior to joining ECVAM. He was pleased that the European Pharmacoepia will hold a peer review panel to review and accept these methods. He was disappointed with the outcome of the peer review panel meeting. He noted that pyrogenicity tests are very expensive and the approval and release of a single product can cost several hundred thousand dollars. The validation studies were set up to assess whether the new tests would outperform the old tests within a set threshold. Only 50% of the samples would be positive in the most sensitive rabbit strain. All of the *in vitro* assays have an accuracy of around 90%. He outlined six points where the BRD had been criticized.

1. ICCVAM said the BRD is deficient due to the limited data for only 10 pharmaceutical substances from the validation studies, which alone cost $6M. The recommendations for additional studies from the peer review panel would cost between $20-40M and they would be a waste of resources because a product-specific validation process would be required for each application. To help contain cost, the tests described in the BRD were designed to emphasize the accuracy of the method to detect pyrogens near the threshold.

2. The peer review panel did not acknowledge the difference in status of the five methodologies. Some methods are used in more than 80 laboratories while others are used infrequently; however, the same criticisms were applied to all of the methods.

3. The BRD recommended that parallel testing be conducted with the RPT. However, parallel testing in rabbits is unnecessary because these studies have been performed for 65 years using a WHO standard as a reference material. The outcome from rabbit testing is so predictable that ethically it is not warranted. Also, in the European Union, it will be impossible for ECVAM to carry out these *in vivo* tests especially as the new methodologies have shown partial concordance.

4. Endotoxins are only tested in the BET assay, and this assay has replaced the RPT for about 90% of substances; the remaining 10% of substances consist of non-endotoxin pyrogen products that interfere with the BET. He asked why the new tests have to meet higher standards than the BET, which has been endorsed for the testing of many pyrogenic products. He noted that no reference non-endotoxin pyrogens are suitable for validation purposes in rabbits and humans; therefore, inclusion of such controls is scientifically impossible.

5. High endotoxin concentrations will be detected accurately in the RPT, BET, or any of the new *in vitro* pyrogenicity assays. Hence, a concentration near 50 pg of endotoxin, which is equivalent to 0.5 EU and is the threshold for rabbits, was chosen for the assays. Additional concentrations of 100 pg and 25 pg were also selected. Even though the assays were challenged at these low concentrations, they were 90% accurate. False positives were due to spikes at half the threshold indicating that the assays are too sensitive.

6. The new assays were evaluated fairly in comparison to the limitations of the existing tests. The rabbit test, which has a number of limitations, has never been properly validated for non-endotoxin pyrogens. The BET does not detect all Gram-positive endotoxins although the new assays have shown some capability for doing so.

In conclusion, the proposed test methods for which data sets have been provided perform better than the BET and RPT. Dr. Hartung proposed that the rabbit assay be replaced with the *in vitro* assays because the RPT cannot match their performance, as reported in the BRD.

SACATM Discussion.
SACATM was asked to address questions regarding the peer review panel's conclusion and recommendations of the draft ICCVAM BRD with regard to its completeness; the panel's identification of errors or omissions; whether ICCVAM's applicable criteria for validation and acceptance of toxicological test methods were addressed; and to provide comments on the draft ICCVAM test methods recommendations, usefulness of the test

methods, the test method protocols, proposed performance standards, as well as proposed additional studies.

Dr. Barile, a lead discussant, said there was no question about the usefulness of pyrogenicity testing and the urgency and importance of validating these tests. In combination, some of these tests will contribute to the reduction of animal usage. One major deficiency of present pyrogenicity testing is that the RPT only detects about 50% of the endotoxins. Some of the proposed *in vitro* tests had false negative responses in the range of 10% while the IL-1 assay had a false negative response of 27%. These false negative responses could be due to consistently higher variability among some donors, which would be a limitation relative to a whole blood human assay. He expressed concern that the IL6 ELISA test, marketed by Novartis, is a proprietary test and he would not recommend approving a method without knowing the experimental details. He agreed with Dr. Hartung that parallel testing in rabbits was unnecessary during development of the testing methodologies. However, a comparison to RPT data is necessary so that a valid concordance or regression analysis between the *in vivo* and *in vitro* methods can be undertaken. He said samples spiked with endotoxin are not representative of real world samples such as a biological vaccine or a solubilized pharmaceutical product. There is no solubility problem associated with the testing of biological vaccines in rabbits, but insolubility is a problem in *in vitro* tests even if the test article is in suspension and this technicality must be addressed. He believes that the cell culture methods are more developed than the whole blood methods for validation purposes. A few additional studies, which address the panel's recommendations, would allow the cell culture pyrogenicity tests to receive validation status.

Dr. McClellan said he was generally pleased with the draft BRD until he heard Dr. Hartung's statement. He did not believe that the BRD is adequate nor can he compliment the peer review panel on its report. He wondered how this difference of opinion would be resolved and asked Dr. Brown to comment.

Dr. Freedman said he was confident that all of SACATM's comments would be taken into account by ICCVAM and, if necessary, ICCVAM could reconvene the expert panel.

Dr. Brown said ECVAM produced a reasonably comprehensive BRD, but the panel was not able to address all of the components of the individual *in vitro* methods because time for discussion was limited. Some of the details were missing or difficult to understand; however, she felt that given more time to discuss these methods, the panel might have been able to provide a stronger recommendation for one or more of the assays. Personally, she felt that the MM6 assay has the greatest potential and several of the other panel members agreed. The most bothersome aspect for the panel was trying to identify the specifics of the validation protocols. She noted that for an *in vitro* assay it is critical to identify every component and every single condition of the assay completely, but this information was not provided, particularly for the MM6 test method. She was impressed with the cell culture methodology, although specifics such as cell passage levels, or how many cells are used in a test were lacking. She felt that the panel did not seem to understand cell culture methodology and its related costs. Consequently, they got side-

tracked in specifics, which hindered them from making progress and reaching conclusions.

Dr. Brown said she does not believe that it is necessary to run *in vivo* assays in parallel with the *in vitro* assays. She is unsure how one can run a regression analysis with one test that is 90% accurate and a second that is 50% accurate. She questioned whether it is necessary to validate an *in vitro* test against an animal test that is not as accurate as the *in vitro* assay itself.

Dr. McClellan said that Dr. Hartung disclosed his own potential biases, concerns, and background. He asked whether Dr Hartung was suggesting that two of the assays should have received more attention and wondered which of the assays Dr. Hartung thought were appropriately validated and whether he might focus the panel toward those assays.

Dr. Stokes said that in the future NICEATM would set aside at least two days for a peer review meeting, so that a panel can fully understand the methodologies before they deliberate on the evaluation questions.

Dr. Qu had some comments on the panel's concern about data transformations. The panel was not sure if the data were transformed and whether or not the use of a "t" test for their analysis was appropriate. She said it is not necessary to use a "t" test even if the data are normal. A non-parametric test such as the permutation test, which does not require transformation, could be used. Dr. Qu noted also that it is important to control for false positives when doing a multiple comparison for several tests. By doing multiple comparisons, it is possible to obtain a statistically significant difference that is not biologically significant. One approach to dealing with this problem is to use a more stringent level of significance.

Dr. Becker welcomed the proposed longer time frame for a peer review meeting. He suggested that it might be useful to convene a meeting with a core panel of validation experts and then have subject-specific experts to address specific assays.