I. LOCATION OF BACKGROUND MATERIALS/PRESENTATIONS AND FREQUENTLY USED ABBREVIATIONS

Background materials and presentations for the SACATM meeting are available on the SACATM meeting web site (*http://ntp-server.niehs.nih.gov* see "Advisory Board & Committees).

CPSC	Consumer Product Safety Commission
DOI	Department of the Interior
DOT	Deptartment of Transportation
EPA	U.S. Environmental Protection Agency
FDA	Food and Drug Administration
ICCVAM	Interagency Coordinating Committee on Alternative Methods
ILS	Integrated Laboratory Systems, Inc.
ECVAM	European Centre for the Validation of Alternative Methods
GHS	Globally Harmonized System of Classification and Labelling of Chemicals
JaCVAM	Japanese Center for the Validation of Alternative Methods
NICEATM	The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NCI	National Cancer Institute
NIH	National Institutes of Health
NLM	National Library of Medicine
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
USDA	U.S. Department of Agriculture

II. ATTENDANCE

SACATM met on December 12, 2005, at the Hilton Alexandria Old Town, 1767 King Street, Alexandria, Virginia 22314.

SACATM Members in Attendance

Daniel Acosta, Ph.D.(University of Cincinnati)
Rodger Curren, Ph.D. (Institute for *In Vitro* Sciences, Inc)
Nancy Flournoy, Ph.D. (University of Missouri - Columbia)
Alan Goldberg, Ph.D. (John Hopkins University)
Sidney Green, Ph.D. (Howard University)
A. Wallace Hayes, Ph.D. (Harvard University)
Nancy Monteiro-Riviere, Ph.D. (North Carolina State University)
Dr. Stephen Safe, Ph.D. (Texas A&M University)
Jacqueline Smith, Ph.D. (Chesapeake Consulting Team)
Carlos Sonnenschein, M.D. (Tufts University)
Martin Stephens, Ph.D. (Humane Society of the U.S.)
Katherine Stitzel, D.V.M.
Peter Theran, V.M.D. (Massachusetts Society for the Prevention of Cruelty to Animals)
Calvin Willhite, Ph.D. (California EPA)

ICCVAM Ex Officio Members in Attendance

George Cushmac, Ph.D. (DOT)
Karen Hamernick, Ph.D. (EPA)
Vera Hudson, M.S. (NLM)
Jodie-Kulpa-Eddy, D.V.M. (USDA)
Alan Poland, M.D. (NCI)
Barnett Rattner, Ph.D. (DOI)
Leonard Schechtman, Ph.D. (FDA)
Margaret Snyder, Ph.D. (NIH)
William Stokes, D.V.M, (NIEHS)
Marilyn Wind, Ph.D. (CPSC)

Ad hoc and Liaison Representatives in Attendance

Frank Barile, Ph.D. (St. John's University)
Richard Becker, Ph.D. (American Chemistry Council)
George DeGeorge, Ph.D. (MB Research Laboratories)

Marilyn Brown, Ph.D. (Charles River Laboratories)
Grantley Charles, Ph.D. (Dow Chemical Company)
Donald Fox, Ph.D. (University of Houston)
Marlies Halder, Ph.D. (ECVAM)

NIEHS/NTP Staff and NTP Contractors in Attendance

Dave Allen, Ph.D. (ILS)
Brad Blackard, M.S.P.H. (ILS)
John Bucher, Ph.D. (NIEHS/NIH)
Jeff Charles, Ph.D. (ILS)
Neepa Choksi, Ph.D. (ILS)
Sally Fields (NIEHS/NIH)
Joseph Haseman, Ph.D. (consultant)
Debbie McCarley (NIEHS/NIH)
Michael Paris (ILS)
Christopher Portier, Ph.D. (NIEHS/NIH)
William Schrader, Ph.D. (NIEHS/NIH)
Judy Strickland, Ph.D. (ILS)
Kristina Thayer, Ph.D. (NIEHS/NIH)
Raymond Tice, Ph.D. (NIEHS/NIH)
Mary Wolfe, Ph.D. (NIEHS/NIH)

Other Federal Staff in Attendance

Abby Jacobs, Ph.D. (FDA)
Richard Canady (FDA)
Steve Hwang (DOT)
Richard McFarland (FDA)
Amy Rispin (EPA)

Public in Attendance

Joel Fisher (International Joint Commission)
James Freeman, Ph.D. (Exxon Mobil)
Hyung Soo Kim (Korean National Institute for Toxicological Research)
Thomas Re (L'Oreal)
Sherry Ward, Ph.D. (consultant)

III. WELCOME, INTRODUCTIONS, AND RECOGNITION OF RETIRING MEMBERS

Dr. Kathy Stitzel, Chair, called the meeting to order at 8:30 a.m. on December 12, 2005, and asked individuals in the room to introduce themselves and provide their affiliation.

Dr. Christopher Portier, Associate Director of the NTP, welcomed SACATM and the ICCVAM and thanked them for attending the meeting. He also distributed certificates of service for SACATM members whose terms expire: Drs. Curren, Goldberg, Green, Hayes, Sonnenschein, and Stitzel. Dr. Schechtman, ICCVAM Chair, and Dr. Stokes, NICEATM Director, welcomed SACATM and the ad hoc attendees.

Dr. Kristina Thayer read the conflict of interest statement for SACATM.

IV. ICCVAM-NICEATM UPDATE

A. Presentation

Dr. William Stokes, NICEATM Director and ICCVAM Executive Director, provided an overview of ICCVAM-NICEATM activities since the October 2004 SACATM meeting. Major topics covered by Dr. Stokes included:

- ICCVAM-NICEATM test method evaluation activities
  - *Ocular Toxicity Test Methods.* Four *In vitro* methods for identifying ocular corrosives and severe irritants were evaluated by an expert panel on January 11-12, 2005: (1) the bovine corneal opacity and permeability (BCOP) test, (2) the hen's egg test – chorion allantoic membrane (HET-CAM), (3) the isolated rabbit eye (IRE), and (4) the isolated chicken eye (ICE). Following the January meeting, additional data were submitted to ICCVAM-NICEATM that was incorporated into the

analyses. A draft addendum to the first expert panel report that included the additional data was discussed by the expert panel via teleconference on September 19, 2005. Background review documents and expert panel reports from the ocular toxicity evaluation are available at *http://iccvam.niehs.nih.gov/ethods/eyeirrit.htm*. The ICCVAM Test Method Report and finalized background review documents are expected to be released in February or March of 2006.

- *Acute Systemic Toxicity Test Methods.* NICEATM and ECVAM are conducting a joint validation study with the overall goal of determining the usefulness and limitations of using cytotoxicity IC50 data as the basis for starting doses for *in vivo* acute toxicity studies. The laboratory component of this effort ended in January 2005 and a peer review panel evaluation is scheduled for May 23, 2006. NICEATM expects to issue the ICCVAM Test Method Report in late 2006.
- *Endocrine Disruptor Test Methods.* NICEATM has initiated standardization studies for the LUMI-CELL™ Estrogen Receptor Screening Assay that should be completed in June 2006. Depending on the results of these studies, an international validation study will be planned to include three laboratories from the U.S, Europe, and Japan. NICEATM is collaborating with ECVAM and JaCVAM in this effort. The list of reference substances proposed for use in the validation studies is currently being revised by the ICCVAM endocrine disruptor working group because some of the substances on the proposed list are very costly or not commercially available.
- *Ocular and Dermal Test Methods for Antimicrobial Disinfectant Products.* A background review document prepared by industry and the Institute for In Vitro Sciences, Inc. is scheduled for submission to NICEATM in January 2006. An expert panel meeting is tentatively scheduled to meet in the late 2006, followed by the release of ICCVAM Test Method Recommendations in early 2007.

- ICCVAM-NICEATM participation in the 5th World Congress on Alternatives and Animal Use in the Life Sciences (Berlin, Germany, August 21-25, 2005)
- ICCVAM-NICEATM collaborations with ECVAM:
  - (1) ICCVAM-NICEATM-ECVAM Workshop: Validation Principles and Approaches for Toxicogenomic-based Methods (December 2003, Ispra, Italy); (2) ECVAM Eye Irritation Expert Meeting (February 8-11, 2005, Ispra, Italy); (3) evaluation of ocular toxicity assays; (4) NICEATM-ECVAM *in vitro* cytotoxicity independent validation study; and (5) ECVAM Acute-Tox Project
- OECD activities:
  - (1) Test Guideline 435: *In Vitro* Membrane Barrier Test Method for Skin Corrosion (adopted 2005), (2) Test Guidance Document 34: The Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (adopted August 2005), and (3) OECD Non-Animal Testing Validation Management Group - Task Force on Endocrine Disrupting Testing and Assessment (3rd meeting scheduled for December 14-15, 2005)
- Establishment of the JaCVAM at the Japanese National Institute of Health Sciences. A dedication ceremony was held December 1, 2005.

B. SACATM Discussion

Dr. DeGeorge commented that the OECD test guidelines are typically not free and asked whether there is a charge for the OECD Test Guideline 435. Dr. Stokes did not believe that a free electronic version is available; however, he noted that the U.S. national coordinator to the OECD would have copies and a hard copy might also be obtained by contacting the OECD office in Washington DC. Dr. Becker asked for additional details on coordination efforts between the EPA and ICCVAM-NICEATM on the validation of estrogen receptor transcription activation assays. Dr. Stokes said ICCVAM-NICEATM is in contact with staff in the EPA Office of Prevention, Pesticides and Toxic Substances and expects to receive documentation for estrogen and androgen receptor binding assays early next year. ICCVAM-

NICEATM intends to present these documents to an expert panel for peer review. Dr. Sonnenshein was concerned about the reproducibility and accuracy of these test methods and asked who would review the reports. Dr. Stokes said an independent panel would review the reports and anyone without a conflict of interest with respect to the methods being discussed could be considered as a potential panel member.

V. ICCVAM NOMINATIONS

A. Presentation

Dr. Raymond Tice, NICEATM Deputy Director, discussed (1) a nomination from the Human Society of the United States (HSUS) to replace the mouse LD50 assay for botulinum toxin potency testing ("botulinum nomination"), and (2) a submission from ECVAM for five *in vitro* pyrogenicity test methods ("pyrogenicity submission").

- ICCVAM-NICEATM received the HSUS botulinum nomination on October 31, 2005. The HSUS nomination included a review of the relevant science, including potential *in vitro* replacements, and several rationales for why ICCVAM-NICEATM should pursue the recommended activities. The HSUS recommended that ICCVAM-NICEATM convene a workshop on this topic and then work with appropriate partners to validate an alternative assay(s). ICCVAM proposes to move forward with the botulinum nomination and has established a biologicals working group that will include an ECVAM liaison to coordinate efforts. The next steps are to publish a Federal Register to (1) request public comment on the nomination, (2) identify other test methods that could be considered, (3) request the submission of data on proposed alternatives for consideration, and (4) solicit the nomination of potential workshop participants. Following the comment period, ICCVAM will finalize its recommendations.
- On June 24, 2005, ECVAM submitted background review documents for five *in vitro* pyrogenicity assays: (1) the human PBMC/IL-6 *in vitro* pyrogen test (PBMC/IL-6), (2) the human whole blood/IL-1 *in vitro* pyrogen test (WB/IL-1) (3) the human whole blood/IL-1 *in vitro* pyrogen test: application of cryopreserved human whole blood (cryo WB/IL-1), (4) the human whole blood/IL-6 *in vitro* pyrogen test (WB/IL-6) , and (5) an alternative *in vitro* pyrogen test using the human monocytoid cell line MONO MAC-6 [MM6] (MM6/IL6). The methods are proposed as replacements to the rabbit pyrogen test and the bacterial endotoxin test (BET). ICCVAM believes the five test methods have potential for pyrogenicity testing, but ECVAM needs to provide additional data prior to formal review of the proposed methods by an expert peer review panel. NICEATM will next publish a Federal Register to (1) request public comment on the submission, (2) request the submission of relevant data, and (4) solicit the nomination of potential expert panelists. Early in 2006 the ICCVAM pyrogenicity working group (PWG) will consider comments made by SACATM and the public to formulate recommendations on future activity. These recommendations will then be forwarded to ICCVAM. If convened, an expert panel will have two major charges: (1) peer review the background review documents for completeness and adequacy, and (2) determine whether the data cited in the background review documents support the draft ICCVAM Test method Recommendations regarding the proposed usefulness, limitations, and validation status of the test methods.

B. SACATM Discussion

Before proceeding with the discussion, Dr. Stitzel asked whether anyone had a conflict of interest (COI) with either the botulinum or pyrogenicity assays discussed. Several individuals noted real or possible COI issues:

- Dr. Green recused himself from any discussions related to the botulinum nomination because he consults for Allergan, the producer of Botox®, on the replacement or development of alternative methods for the LD_{50} assay.

- Dr. Stephens asked whether he is excluded from discussing the botulinum nomination since it was submitted by his organization.
- Dr. Goldberg said he met with Allergan one time at the request of the HSUS. He was not acting as a paid consultant in this activity.
- Dr. Brown commented that her company, Charles River Laboratories, has a business unit that produces the Limulus Amebocyte Lysate (LAL) test. She may have a COI issue as the LAL test is a potential replacement for the rabbit pyrogen test.
- Dr. DeGeorge said his lab has conducted a version of the BET and *in vivo* pyrogenicity testing; however, his lab does not currently conduct either of these tests.

Dr. Hartung said that three of the five test methods for pyrogenicity discussed in the ECVAM submission are based on patents he holds from 1995-1996. These patents are being pursued by a non-profit technology transfer center at the University of Konstanz. Although he has never received any personal payments as a result of these patents, some of his academic research has been supported by license fees from the patents.

Drs. Theran and Goldberg served as lead discussants for this agenda topic. Dr. Theran is very supportive of a workshop to pursue alternative methods to replace the mouse LD50 assay because he believes the mouse assay is imperfect and costly to animals. He is also very supportive of pursuing validation efforts on the alternatives to the *in vivo* pyrogenicity tests discussed in the ECVAM submission. He made a general comment expressing his concern that validation efforts are not necessarily being appropriately prioritized to address older methods that yield the highest costs for animals. For example, there is discussion of pursing new research methods (e.g., toxicogenomics) while certain *in vivo* tests, such as the Draize eye test and the botulinum protocol, cause a level of pain and distress that would likely not be approved now without modifications to the protocols. Dr. Theran believes the priority should be to validate methods for *in vivo* tests that present the greatest costs to animals rather than towards new research tools.

Dr. Goldberg thought both packages received by ICCVAM-NICEATM were well prepared and presented. However, while he strongly supports pursuing the ECVAM pyrogenicity submission, he has reservations about supporting the HSUS botulinum nomination. First, he is concerned that the proposed alternative methods will not mimic the *in vivo* events required to assess the potency of botulinum. There are two steps that need to occur *in vivo* to assess potency: (1) botulinum has to access the pre-synaptic neuron and (2) interfere with the release of acetylcholine. While the proposed alternative methods can detect interference, he questions whether they can appropriately model access. Second, the botulinum LD50 assay only impacts two or three companies and it is unclear how many mice are being used in the potency test. In contrast, the *in vivo* pyrogenicity tests impact numerous companies and involve a substantial number of horseshoe crabs and rabbits, probably at least 250,000. If the LD50 test used by companies to assess botulinum potency is conducted in the classic sense then many mice would be involved; however, if companies are using the updated LD50 approach then far fewer animals would be used. For this reason he suggests that ICCVAM contact the companies to find out what type of LD50 protocol is being used and whether they are pursuing alternatives. He is not convinced that ICCVAM should be pursing this effort because it may be best addressed by the companies involved. Dr. Stokes thought this was an excellent suggestion and said contacting the companies would be one of the first steps undertaken to evaluate the nomination.

Dr. Willhite asked Dr. Goldberg for clarification on whether he thinks the proposed assays are essentially efficacy assays and, if so, stated that this is separate from the issue of bioavailability. Dr. Goldberg said this is basically true. Dr. Willhite asked if it would be useful to also explore approaches to assess topical or systemic bioavailability. Dr. Goldberg said it is not that simple as botulinum is injected and then must reach the active site; this process is very variable because it is a biologically prepared material. Dr.

Richard McFarland from the FDA Center for Biologics Evaluation and Research said the assay is more than an efficacy test – it is also a neurotoxicity test. He commented on a case in Florida where non-pharmaceutical grade Botox® was administered and resulted in significant toxicity. He also noted that FDA is aware of other alternative methods that should be included in the workshop, if it occurs, including one being developed within the Center for Biologics Evaluation.

Dr. Golberg noted that the current botulinum unit is equivalent to the LD50 and asked how difficult it would be to re-define the standard unit of Botox®. Dr. McFarland could not fully answer because he didn't want to reveal privileged information. However, he said it is possible and would probably take a small validation study for a new method to change the indicator from the LD50.

Dr. DeGeorge said his company has experience with a modified version of the HET-CAM assay that appears to be very sensitive for various toxins. He suggested this assay would be useful from both the efficacy and safety perspectives and should be considered if a workshop is convened. He noted that his company has approached Allergan on this issue, but has not yet engaged in discussions with them. Dr. DeGeorge also commented that 4 of the 5 proposed substitutes for the existing *in vivo* pyrogenicity tests involve the use of human blood. Use of these assays would require a biosafety level 2 or higher containment tissue culture facility that would greatly increase the costs of conducting them. He also thought the list of substances used for determining test method accuracy is too short and narrowly focused and suggested expanding the list. Dr. Stitzel clarified that Dr. DeGeorge is suggesting that the modified HET-CAM assay is relevant to both the botulinum and pyrogenicity topics.

Dr. Barile asked about the process by which ICCVAM-NICEATM considers nomination and submissions. Dr. Tice responded that ICCVAM-NICEATM receives two general types of packages: nominations and submissions. The difference between the two is that a nomination asks ICCVAM-NICEATM to explore alternative methods for which sufficient validation data are not available while a submission package contains validation study data and asks ICCVAM-NICEATM to evaluate the validation status of the proposed test methods. Typically, NICEATM organizes workshops for nominations and expert panel meetings for submissions. Once a submission package is received, NICEATM conducts a pre-screen evaluation to assess whether the information is complete and adequate enough to warrant an expert panel meeting. This pre-screen evaluation is then forwarded to an appropriate ICCVAM working group composed of scientists from ICCVAM member agencies who have experience with the assay under review. The ICCVAM working group then determines whether the submission package is complete or whether additional information is needed, as was the case for the ECVAM submission. In this instance, the additional, requested information related to how the proposed test methods would fit into the FDA regulatory paradigm on pyrogenicity. The working group recommendations then go before ICCVAM and are released for public comment in a Federal Register notice. In addition, ICCVAM asks for the submission of relevant data and the names of other assays that should also be considered. However, Dr. Tice said consideration of other assays is more likely to occur as part of a nomination-workshop activity rather than an expert panel evaluation. In regard to the pyrogenicity submission, the ICCVAM PWG and ECVAM will be meeting in the near future to discuss the additional, requested information. The outcome of this discussion, along with public comments and SACATM comments, will then be reviewed by ICCVAM as it makes a conclusion on the appropriate activity and priority of that activity. At future expert panel meetings, panelists will (1) review the background review documents for a test method, (2) review the draft ICCVAM Test Method Recommendations on the uses and limitations of the method(s), (3) determine whether the protocol has been sufficiently optimized, and (4) determine whether performance standards need to be developed. Dr. Portier clarified that additional methods suggested in the context of a submission could be considered nominations.

Dr. Fox said he understands that the botulinum test is currently only a potency evaluation. If so, then he agrees with Dr. Theran that an *in vitro* test to assess potency is feasible. Efficacy and bioavailability are different issues. Dr. Tice clarified that Dr. Fox is endorsing a workshop on the HSUS nomination. Dr. Curren asked Dr. McFarland whether a workshop would assist or hinder FDA in its efforts on this issue. While he could not speak for the FDA center that currently regulates Botox®, Dr. McFarland thinks the timing of a workshop would factor into it utility. He also thought a workshop would be most successful if the companies that produce Botox® are able to participate and discuss what they might be working on. Dr. Abigail Jacobs from the FDA Center for Drug Evaluation and Research said it is her understanding that the FDA group working on this issue supports a workshop and has been active intramurally working on alternatives.

Dr. Stephens commented that the HSUS nomination is not meant to portray this issue as being simple, but rather to propose assessing the science to determine what is achievable. He believes the botulinum issue is an opportunity to be proactive in terms of identifying potential "low hanging fruit." He was very pleased to hear that ICCVAM is establishing a biologicals working group to evaluate not only botulinum, but also other biologicals since these tests result in painful Category E experiments. Dr. Stephens also noted that this effort may have implications for bioterrorism and diagnosis. Dr. Hartung said the ECVAM biologicals task force intends to coordinate with botulinum producers to develop a replacement strategy and that any workshop would/should be a joint ICCVAM-NICEATM-ECVAM venture. Dr. Goldberg emphasized the need to find out what the manufacturers are doing as to not direct valuable resources towards an activity that may already be being undertaken. Dr. Becker noted the limited applicability and wondered whether a more directed effort to include dialogue with the appropriate regulatory agencies and companies would be a better first step rather than organizing a workshop.

Dr. Stitzel asked if any members of SACATM disagreed with ICCVAM pursuing the botulinum issue in a step-wise fashion to collect other information and potentially including a workshop as a high priority. No SACATM members disagreed.

Dr. Green asked for more detail on what additional information was requested from ECVAM regarding the pyrogenicity submission. Dr. Tice replied that ICCVAM had several comments on the background review document related to references and layout. In addition, the ICCVAM PWG asked for additional detail on the use and limitations of the assay with respect to the U.S. FDA. More specifically, the background review document indicated that the proposed test methods are applicable to endotoxin and non-endotoxin pyrogenic responses, but no data are provided to support non-endotoxin uses. Dr. Hartung will meet with the ICCVAM PWG in January to clarify these issues.

Dr. Green questioned the breadth and representativeness of the 13 substances used to attempt to validate the 5 pyrogenicity assays. Dr. Tice said the expert panel would be asked to comment about the representativeness of the proposed substances for validation studies. This is a valid question as pyrogenicity testing also applies to devices and it is unclear whether the proposed methods are applicable to that component of testing. The intent of the pre-screen conducted by NICEATM is not to evaluate the science, but to address whether sufficient information is presented in the submission to allow a scientific evaluation. Dr. Hartung noted that ECVAM is not addressing substances per se, but the contamination of various drugs, so the substances are best considered as matrices for endotoxins, and were selected to challenge endotoxin determination. Dr. Green also asked if the pyrogenicity testing issue is applicable to other U.S. regulatory agencies besides the FDA. He wondered whether this effort should be given a high priority if it does not have broad applicability to regulatory agencies. Dr. Tice said broad applicability is just one factor used to determine priority. Other factors include the number and types of animals used, the extent to which animals undergo pain and suffering, and the ease of finding a replacement. ICCVAM-NICEATM also considers resources, but so far no nomination or submission has been rejected due to a lack of resources. Dr. Green suggested looking at the costs associated with the replacement

assays – the proposed assays seem rather intensive and potentially expensive and this might be an issue for industry adopting them. Dr. Tice said ICCVAM-NICEATM have requested more information on costs. In addition, ICCVAM-NICEATM typically queries various organizations that conduct testing to get realistic cost estimates. With regard to cost, Dr. Hartung said several contract companies are offering the proposed test methods at 1/5th the cost of the rabbit test.

Dr. Stitzel asked if any members of SACATM disagreed that pursuing the pyrogenicity submission should be a high priority for ICCVAM. No SACATM members disagreed.

VI. PERFORMANCE CHARACTERISTICS OF THE *IN VIVO* OCULAR IRRITANCY TEST

A. Presentation

Dr. Haseman, a consultant retired from the NIEHS, presented on the evaluation of the under- and over-classification likelihood of the *in vivo* rabbit ocular irritation test method. The study objectives were to evaluate the likelihood of (1) under-classifying an ocular corrosive or severely irritating substance as a nonsevere irritant/ nonirritant in the current rabbit eye irritation test, and (2) over-classifying an ocular nonsevere irritant or nonirritant as a corrosive or severely irritating substance. The estimated overall under- and over-classification likelihoods are summarized in the table below.

GHS	Original Analysis			Modified Springer Analysis[d]
	Calculation 1[a]	Calculation 2[b]	Calculation 3[c]	
Under-classification	**4.30%**	**13.24%**	**11.21%**	**13.40%**
Over-classification	**0.88%**	**0.74%**	**-**	**2.65%**

[a] *Homogeneity of response* – assumed that all substances have the same pattern of rabbit response within an irritancy category.
[b] *Heterogeneity of response* - assumed that substances have a different pattern of rabbit response within an irritancy category
[c] *Homogeneity/Heterogeneity of response combination* - assumed that category 1 substances have similar patterns of rabbit response within 3 categories (strong, moderate, or weak responders)
[d] *Modified Springer analysis* assumes a mixture of three binomial distributions and allows heterogeneity among irritants by assuming that the irritants are a mix of two homogeneous subcategories: strong and weak irritants

Major summary points are:

- Since the Category 1 irritants clearly showed heterogeneity of response, the most reasonable estimate of the under-classification likelihood for the current rabbit eye test method is between 11-14%.
- For the subset analyses:
  - Criterion 1 (severity classification based on persistence) was shown to have a higher under-classification likelihood than Criteria 2-4 (severity classification based on severity).
  - The under-classification likelihood for liquids/gels was slightly higher than for solids, but the difference was not statistically significant.
  - Under-classification likelihoods appear to vary by chemical class.
- The estimated over-classification likelihood for the *in vivo* rabbit eye test method is less than 3%.

B. SACATM Discussion

Dr. Flournoy asked for clarification on the method for data sampling. Dr. Haseman replied that the estimated under- and over-classification rates were based on a sequential testing strategy, which used up to three rabbits in which the probability of response for a given rabbit was determined by the observed distribution of responses within that irritancy category. Dr. Flournoy suggested clarifying the language in the methods section of the report. She also asked how the analysis would be used. Dr. Stokes said this type of analysis might be used to compare the relative performance of any new *in vitro* test to the existing *in vivo* test. For example, this information might be used to determine whether a proposed *in vitro* test could be used as a screen in a tiered testing strategy or as a complete replacement for an existing test. However, the values presented are not intended to serve as "bright lines" against which *in vitro* methods are compared. Other information, such as reversibility of the effect, is also considered. Dr. Flournoy suggested that variability estimates be calculated for the over- and under-estimates provided in the analysis since they are point estimates. Dr. Haseman thought this could be done. Dr. Haseman noted that prediction rate estimates are based primarily on the variability among animals within a test and that data on reproducibility are not available. In addition, the existing data are not sufficient to evaluate the extent to which results in rabbits predict human response. His analysis essentially evaluates how well the rabbit predicts rabbit (i.e., variability within a test).

Dr. Stephens supported the effort to evaluate performance characteristics and understood the limitations of the database. He thought that although additional historical data on the variability and accuracy of the rabbit test likely exist, this data would be very difficult to obtain. He thought the absence of reproducibility data, both within and across laboratories, is a major drawback. For this reason, the estimates of variability provided are best considered minimal estimates of variability. In addition, variability is likely higher when the irritancy response is not extreme. He shared Dr. Haseman's concern on the circularity of the analysis in the sense that rabbit data are being used to both classify and assess performance characteristics and appreciated the inclusion of the Springer modified analysis to help address this issue. He questioned what effect the prevalence of substances in different irritancy categories might have on the analysis. Dr. Haseman did not believe prevalence would influence the under- and over-classification estimates, but it could influence the performance of the test itself in the sense of understanding how likely a positive outcome represents a "true" irritant. With respect to prevalence, Dr. Portier commented that it wouldn't have any impact if the same compounds tested *in vivo* are used to validate an *in vitro* test.

Dr. Hayes was concerned that language in the report strongly implies that the performance estimates would be used as benchmarks when comparing the *in vitro* and *in vivo* tests. Dr. Stokes said they do not have to match exactly, but any alternative method must provide for equivalent or better protection of human health. Dr. Stitzel said this is an important point because within- and between-laboratory variability are considered when validating an alternative test method, but have not been adequately considered in evaluating many *in vivo* tests. Dr. Portier said this analysis identifies the limitations of comparing *in vitro* to *in vivo* data.

Dr. Fox had some concerns about the 1A and 1B sub-classifications because long-term, acute, and persistent effects are not incorporated in these categories. He thought attention should be directed to understanding any under classification occurring in criterion 1 with respect to long-term effects and subtle differences between 1A and 1B. Dr. Tice replied that the separation between 1A and 1B is an internal classification to aid Dr. Haseman in the analysis. There are two ways a compound can be classified as GHS category 1. The first is based on 1 of 3 animals responding by either having a corneal opacity score of 4 at anytime or an animal that has a persistent lesion out to day 21. The second is based on severity in terms of having an average corneal opacity score of 3 in at least 2 of 3 animals across 24, 48, and 72 hours or an iritis score of 1.5 across the 3 days. NICEATM established 1A and 1B sub-classifications

because they were believed important to distinguish between these 2 classification schemes. This led NICEATM to evaluate GHS category 1 classifications made on persistence versus severity with the expectation that classifications made only on persistence would be more heterogeneous within animals. For severe responses, corneal opacity score of 4, there is much less variability. Dr. Tice said NICEATM also tried to assess correct identification of persistent effects although the database was not sufficient to really evaluate this issue. In at least one case, the ability to identify severity based on persistence only yielded the greatest misclassification. Dr. Fox asked that such statements be included in the final performance characteristics report.

Dr. Willhite asked whether there are any published studies that have calculated correlation coefficients between the Draize test and the outcome of a proposed *in vitro* test, such as the BCOP. Dr. Haseman expects to work with NICEATM on this issue in the future. Dr. DeGeorge said he believes Merck has data on the correlation between the BCOP and *in vivo* data, probably rabbit, and the measures of accuracy were in the mid-80s. In his experience, other alternatives also positively predict *in vivo* data with correlations in the high-80s to low 90s. Dr. DeGeorge said the accuracy of the BCOP goes up into the 90s compared to either rabbit or human data (based on a few hundred comparisons) when histology is added. The weak point is negative predictivity. With respect to over and under predictions, his company has reviewed rabbit data from 1972-1992 for about 2400 assays conducted with essentially the same staff and found that the rabbit response appears to be 25-40% more sensitive than the human eye when extreme responses (non-irritancy or corrosion) are excluded. His company is in the process of publishing this analysis.

Dr. Curren asked whether use of the language “irritant” vs. “non-irritant” meant “category 1 irritant” versus “non-category 1 irritant.” Dr. Haseman said this was true for the Springer approach whereas in the first analysis presented, an effort was made to distinguish also between category 1 severe irritants, category 2A moderate irritants, category 2B mild irritants, and non-irritants. Dr. Curren expressed his concern that this analysis only addresses within experiment variability. He asked whether different terminology could be used to characterize the report, such as “maximum obtainable performance” because he does not feel that the caveats to the analysis come through in the title and summary. Dr. Haseman said the final report will give these issues more emphasis. In addition, Dr. Curren said it is his understanding that graders of lesions for the Draize test are not always blinded as to what compound the rabbit was exposed and this non-blinding could reduce within experiment variability. Dr. Curren suggested NICEATM look at the studies submitted by the Cosmetic, Toiletry, and Fragrance Association separately because they were conducted recently, are of high quality, and were conducted in one laboratory using a randomized block design. Dr. Curren asked whether the performance characteristics report would be peer-reviewed in addition to being presented to SACATM for comment prior to its release. Dr. Tice replied that the report would be submitted to a journal publication and would be peer-reviewed as part of that process.

Dr. Stitzel thought the analysis is very useful in a number of ways, such as for identifying groups of chemicals with more or less variability. It is extremely useful to know which types of substances yield consistent and variable results in the *in vivo* studies so that these factors can be considered when evaluating performance of any alternative test method. Dr. Haseman agreed, but noted that conclusions about *in vivo* variability in the current analysis are being based on ~ 6 rabbits and ideally variability estimates would be based on multiple independent studies. Dr. Flournoy asked whether alternative methods to model heterogeneity were considered. Dr. Haseman said two methods were utilized, but he would be happy to discuss additional methods if she has any suggestions. Dr. DeGeorge noted a shift in the chemical dataset after the early 90s when cosmetic companies began using *in vitro* tests to evaluate ocular irritancy because they did not have to fulfill prescribed regulatory requirements. This practice reduced the types of chemicals being tested in the Draize rabbit test (i.e. more pesticide oriented). His company is only doing Draize when required and is trying to steer people to the alternatives.

Dr. Stitzel summarized the discussion by saying SACATM believes that this type of analysis is very useful despite (1) the caveats raised regarding measures of variability inherent with the point estimates of performance and (2) concern about using the values as absolute benchmarks against which alternatives would be compared. No members of SACATM disagreed with this summary.

VII. ICCVAM EVALUATION OF *IN VITRO* METHODS TO IDENTIFY OCULAR CORROSIVES AND SEVERE IRRITANTS: REPORT FROM THE EXPERT PANEL

A. Presentation

Dr. James Freeman, Head of the Downstream / Upstream Section at ExxonMobil Biomedical Sciences, Inc., summarized the finding of the ocular expert panel for four *in vitro* test methods evaluated for detecting ocular corrosives and severe irritants: (1) the Bovine Corneal Opacity and Permeability (BCOP) assay, (2) the Isolated Chicken Eye (ICE) assay, (3) the Isolated Rabbit Eye (IRE) assay, and (4) the Hen's Egg Test – Chorion Allantoic Membrane (HET-CAM). For each assay, he summarized the statistics for accuracy,[1] sensitivity, specificity, positive predictivity, negative predictivity, false positive rate, false negative rate, and the total number of *in vivo* severe and nonsevere substances used for the analysis. He also presented some of these statistics specific to the ocular hazard classification system used (GHS, EPA or EU), chemical class of the substance, and properties of interest for the substances (e.g., liquid, pH, etc). In brief, the expert panel conclusions regarding use of the test methods in their current forms were:

- *BCOP*[2] - Useful to identify ocular corrosives and severe irritants, with the following exceptions: alcohols, ketones, and solids. In addition, histopathological examination must be added, unless the substance is from a class of materials known to be accurately predicted using only opacity and permeability in the BCOP assay. The panel also concluded that there is a need to confirm that the BCOP identifies substances known to cause serious eye injury in humans.
- *ICE* - Useful to identify ocular corrosives/severe irritants in a tiered-testing strategy, with the following limitations:
  - Alcohols tend to be over-predicted, surfactants tend to be under-predicted, and solids and insoluble substances may be problematic as they may not come in adequate contact with the corneal surface (leading to under-prediction).
  - The low overall false positive rate (8-10%) means that the ICE test can be used at the present time to screen for severe eye irritants.
  - Given the high false positive rate (50%) calculated for a small number of alcohols (n=10), caution should be observed when evaluating ICE test results with these types of substances.
- *IRE* – Although more data are needed, the test method appears to be useful in a tiered-testing strategy for identification of severe irritants/corrosives, with the following caveats:
  - Decision for use of the test method in a tiered-testing strategy requires a larger set of data (increased n) to corroborate the accuracy results and provide a reliability assessment.
  - The addition of fluorescein penetration and epithelial integrity measures may improve accuracy, but limited data are available for analysis with this protocol (n=36 substances for GHS and EPA analyses).

[1] Minority opinion: Drs. Stephens and Theran believe that the term "accuracy" is inappropriately used, and that it is more appropriate to use the term "consistency with *in vivo* data" when comparing test results.

[2] Minority opinion: Dr. Freeman expressed no opinion as to whether the BCOP assay had met the validation criteria as set forth in the ICCVAM Submission Guidelines (2003). This is because the panel did not reach a conclusion on question of whether these validation criteria had been met.

- *HET-CAM* - Based on the revised analysis, the IS(B) analysis method (according to Kalweit) is not sufficiently predictive of ocular corrosives and severe irritants.

- For each test method, the expert panel proposed modifications to the current protocols and provided a variety of recommendations for optimization and validation studies. The panel also proposed a list of reference substances.

B. SACATM Discussion

Dr. Willhite noted that correlation coefficients between the *in vitro* test methods and the Draize rabbit test were not presented. He wondered whether some of the more contentious expert panel discussions were a result of differences in interpretation as to what "adequate accuracy and reliability" and "accurately predicted" mean. Dr. Willhite also commented on differences in standard operating procedures used by different laboratories. He believed that these assays should be able to move forward. Dr. Stitzel said although the expert panel made suggestions on how to improve the test methods, it is not clear whether they absolutely need to be done prior to use of the methods because improvements could be suggested for any assay. Dr. Monteiro-Riviere asked if there are any data to support the range of tissue freshness determined to be acceptable for use (fresh to up to 12 hours); she was concerned that variability in freshness could contribute to apparent inconsistencies. Dr. Freeman was not aware of any data to support the range and believed it was based on anecdotal experience, which is why the expert panel recommended refining this guidance.

Dr. Acosta asked whether there was a consensus among industry and regulatory agencies on the status of these methods. Dr. Freeman said EPA originally requested the validation effort and some companies are already using one or more of these tests to the extent that they can. Other companies are waiting to see whether the methods are approved by regulatory agencies before they use them routinely. Dr. Acosta asked if the European view of these assays is similar to the U.S. view. Dr. Hartung said the test methods have already been accepted for the classification of severe corrosive agents in Europe. In Europe they are in use based on the experience of different regulatory agencies rather than as a result of a formal validation exercise. He agreed with the conclusions of the expert panel. Dr. Stokes elaborated that in Europe, positive results could be used for labeling purposes, but negative results would still need to be followed by animal testing.

Dr. Smith thought the panel did an outstanding job. She also thought that additional mechanistic studies would be valuable to better understand the chemical classes that produce false positive or false negative responses.

With respect to the type I accuracy table for the BCOP, Dr. Portier wondered whether companies would be satisfied with a 20% false positive rate and whether they have the option to follow up a positive *in vitro* result with an animal test. Dr. Tice replied that the BCOP was deemed useful with the exception of alcohols, ketones, and solids. If these classes of compounds are excluded from the analysis, then the false positive rate for severe irritants is 12% and the false negative rate is 0%, which are about the same as for the Draize rabbit test. Dr. Stitzel believed industry would accept a false positive rate of between 10 – 20%. Dr. Hayes said a positive *in vitro* result might be confirmed with an animal study depending on the intended use for the substance. Dr. Portier asked whether the false positive rate was low enough to really impact animal use and suffering. Dr. Hayes understood the point, but said the *in vitro* test methods, if validated, would be the only acceptable approach for use in the European Union.

Dr. Goldberg commented that specificity and sensitivity analyses are not really designed to address the current questions and wondered whether they are the appropriate statistical approaches to use. Dr. Portier said specificity and sensitivity are the right analyses if you accept one of the assays as truth; however, other methods could be used if no assay is accepted as truth. Dr. Fox asked about whether the change in

the HET-CAM data before and after the additional information was incorporated (it went from 0 up to 34% for false negatives) was a result of the Zebet data. He noted that that false negative rate was originally based on 0/12 and increased to 12/40 with the additional data which means that the new data had a false negative rate of 12/28, or almost 50%. Dr. Tice said it was based mostly on Zebet data, but also some from Gautheron. Also, the first data set was limited to formulations. The reanalysis did not reveal any distinct chemical class driving the increased false positive results. Dr. Stitzel asked what activities could occur to improve these methods for characterizing alcohols, ketones, and solids. She would not want to re-do the validation. Dr. Stokes said ICCVAM would recommend a certain protocol(s) be used that might be modified to accommodate other categories.

Dr. Stizel summarized by asking whether SACATM agreed with the reports. No member of SACATM disagreed.

VIII. ICCVAM-NICEATM-ECVAM WORKSHOP ON VALIDATION PRINCIPLES AND APPROACHES FOR TOXICOGENOMIC-BASED METHODS

A. Presentation

Dr. Schechtman presented a summary of a workshop to discuss toxicogenomic (TG) methods held at ECVAM in Ispra, Italy on December 11-12, 2003. The workshop report is being published in Environmental Health Perspectives.[3] He also discussed the regulatory acceptance of TG data and ICCVAM's role in "omics."

- The FDA is soliciting omics data and released a document titled "FDA Guidance on Pharmacogenomic (PG) Data Submissions" in March 2005. However, at this point TG data are only likely to be used as an adjunct to traditional toxicology tests.
- Some of the unresolved issues surrounding TG-based methods include: standardization, validation, data interpretation and significance, extrapolation of results across species/populations, and data transmission, processing, and storage.
- ICCVAM's role in this area could be to (1) share knowledge and develop guidelines relating to the conduct and evaluation of validation studies for alternative methods for regulatory acceptance purposes and (2) facilitate adoption of any new TG-based test methods.
- The December 11-12, 2003 workshop is the first in a series of planned workshops to address validation principles relevant to TG-based methods. The format of the workshop included three breakout groups to discuss the biological, technical, and regulatory aspects of TG-based methods.

B. SACATM Discussion

Dr. Safe said this is a great technology and is apparently being used with some utility for predicting the success of treatments for certain cancers; however, it is not ready for use to predict toxicity for chemicals, especially in *in vitro* assays. There are simply too many sources of variability. He believed the focus on TG data obtained from *in vivo* studies should be on characterizing signature events, but he has great reservations for applying TG data obtained from *in vitro* systems. Dr. Schechtman acknowledged the difficulties and said pursuing this issue would likely occur in a step wise fashion, from *in vivo* to *ex vivo* and moving slowly to *in vitro* approaches. He noted that current validation discussions for microarray technologies have little to do with validation from the perspective of regulatory acceptance. Dr. Safe thought this process would take years and likely decades. Dr. Sonnenschein agreed with Dr. Safe and added that NIH study sections' review of grants that include microarray techniques are not looked upon favorably because of inconsistency in results and the immense and essentially uncontrollable amount of

[3] Validation of Toxicogenomics-Based Test Systems: ECVAM-ICCVAM/NICEATM Considerations for Regulatory Use, Corvi et al. Environ Health Perspect (in press as of December 2005)

variation. He did not believe microarray techniques are being used successfully for diagnostic purposes. Dr. Curren thought that some applications of these technologies could be used in the near term, for example, as early indicators of endpoints used in regulatory toxicology. In this case, the predictions would likely be based on the same animal and tissues. Dr. Curren also wondered the extent of involvement by the National Institute for Standards and Technology in this area. He also said a reoccurring issue is the manufacturer's responsibility to develop an alternative approach, such as ensuring quality assurance and control

Dr. Barile thought these methodologies are probably at the same point in development as other *in vitro* alternatives were 15 – 20 years ago. He wondered if any laboratory is proposing the use of these as predictive tools. Dr. Portier said there are efforts underway to develop datasets to help address this question. For example, the NIEHS has one such effort looking at acute liver toxicity and genomics patterns. Dr. Portier also noted the tremendous variation inherent in these approaches despite stringent adherence to Good Laboratory Practice. NTP has found an order of magnitude variation between animals in gene response. His believes that it is questionable as to whether TG will ever be of great utility because of the volatility involved with measuring something at the molecular level that can be changing from second to second. He thinks SACATM is essentially suggesting that ICCVAM-NICEATM treat this topic as a submission issue where a method is proposed rather than as a nomination that ICCVAM-NICEATM pursues.

Dr. Fox said this brings up the question of what constitutes a toxic response. Overall, he thinks these techniques are useful as tools used to generate mechanistic questions for individual laboratories to pursue, but questions the application to screen chemicals with unknown biological effects. Also, in his experience "omics" techniques are only useful *in vivo* at this point. Dr. Schechtman said he recognizes the steep learning curve. Dr. Fox said it could be more than a learning issue. Dr. Schechtman said one of the reasons why regulatory agencies are soliciting TG data from industry is to get a better idea for where the problems are. Dr. Fox asked if the goal is to get early predictive indicators of toxicity. Dr. Schechtman said this would certainly be one goal. Dr. Hartung added that validation does not necessarily mean to propose a method, but can also mean to challenge a method. Establishing that there is a long way to go for TG methods is necessary to confront the proposals that are now being put forth.

Dr. Willhite recommended that ICCVAM-NICEATM not commit to further effort in this area in the near future. Dr. Stitzel said she believed there is consensus among SACATM on this point.

IX. ECVAM UPDATE

A. Presentation

Dr. Hartung updated SACATM on ECVAM activities, including discussion of:

- The validation process, validation activities, workshops and reports.
- The rational and components of Intelligent Testing Strategies
- Progress reports from key areas/actions including:
    - skin irritation validation study
    - evaluation of methods to detect eye irritation
    - endocrine disruptor validation
    - *in vitro* cell transformation assay as a potential alternative to 2-year cancer bioassay
    - mutagenicity/genotoxicity
    - biologicals
    - ecotoxicology
    - toxicokinetics

B. SACATM Discussion

Dr. Acosta asked how much money was being spent on a yearly basis to support these activities. Dr. Hartung said the work is shared by various parts of the European Commission. The EU Directorate General on Research funds about 16 million euros per year for research projects, ECVAM can apply for these funds. The ECVAM budget is ~8-9 million euros per year. Thus, the total is ~25 million euros per year. Dr. Acosta asked is the NIH, FDA and EPA representatives could present similar numbers. Dr. Portier replied that this is difficult because NIH and other funding in support of the alternatives are often considered basic research. Approximately 3 million dollars per year is used to support ICCVAM, NICEATM, and SACATM. If NIH, NIEHS and NTP monies directed towards high-throughput screens (HTS) are included, then the yearly amount could be as high as 100-200 million although the HTS assays are not being considered for validation in terms of regulatory acceptance at this time.

X. OVERVIEW OF THE MAY 2005 OCULAR TOXICITY SCIENTIFIC SYMPOSIA

A. Presentation

Dr. Stokes presented a summary of the ocular toxicity scientific symposia that was held May 11-13, 2005, in Bethesda, MD:

- The first symposium, "Mechanisms of Chemically-Induced Ocular Injury and Recovery," took place on May 11-12, 2005. The goal of this symposium was to review the state-of-the-science and understanding of the pathophysiology and mechanisms of chemically-induced ocular injury and recovery in order to advance the development of test systems that are (1) necessary to meet regulatory testing requirements, and (2) that will provide for the protection of human health while reducing, refining, and/or replacing the use of animals. Below is a brief overview of the topics discussed by the three panels that met as part of this symposium.
  - *Panel 1 – Mechanisms and Biomarkers of Ocular Injury*. This panel discussed: known mechanism and modes of action, knowledge gaps, research initiatives to address the knowledge gaps, and *in vivo* biomarkers to predict severity, reversibility of lesions, or delayed response.
  - *Panel 2 – In vitro Models of Ocular Injury and Recovery*. This panel discussed: additional biomarkers to include in *in vitro* test systems, *in vitro* systems and biomarkers needed to predict ocular injury and reversibility of damage, and knowledge gaps in differences between *in vitro* and *in vivo* biomarkers.
  - *Panel 3 – In vivo Quantitative Objective Endpoints to Support Development and Validation of Predictive In vitro Models*. This panel discussed: endpoints/biomarkers to include in current *in vivo* rabbit eye test to support the development and validation of alternative models, endpoints/biomarkers to evaluate in human chemical disease to support the development and validation of alternative models, and knowledge gaps.
- The second symposium, "Minimizing Pain and Distress in Ocular Safety Testing," took place on May 13, 2005. Participants at the symposium discussed:
  - Strategies to reduce pain and distress arising from initial application of test articles and post-application tissue injury.
  - Possible interference with irritancy/corrosivity testing as a result of using pre-treatment analgesics or post-treatment use of analgesics and anesthetics.
  - Research needs
  - Biomarkers that can serve as early humane endpoints for ocular injury

B. SACATM Discussion

Dr. Stephens asked if there are any plans for outreach to the Institutional Animal Care and Use Committee (IUCAC) and other decision-making bodies and companies that conduct these types of studies on a routine basis. Dr. Stokes said publications based on the symposia will be widely disseminated and he thinks it's very likely that the ICCVAM Ocular Working Group will be making recommendations for current testing guidelines. Dr. Green asked if there would be any attempt to prioritize the large number of recommendations that came from the workshop. Dr. Stokes replied that some of the near-term recommendations, such as using topical anesthetics and systemic analgesics, could be implemented within the next 6 months. He believes the bigger challenge is the research and development efforts directed towards the *in vitro* systems. The hope is that the symposia recommendations will be considered by funding organizations as well at those who carry out this research. Dr. Stitzel commented that the recommendations to incorporate histopathology and fluorescence standing slit lamps on rabbit eyes have been made repeatedly and she was not sure anything would happen without ICCVAM activity.

Dr. Hayes assumed the role of chair since Dr. Stitzel had to leave.

Dr. Acosta said it was important when these reports are published that past efforts in this area be reported. In this case, many of the same recommendations were made at meetings held years ago. With respect to biomarkers, Dr. Curren wondered how the *in vivo* information would be collected. He thought accomplishing the *in vitro* recommendations would be easier and could potentially have more short-term impacts. He suggested that the list of recommendations be prioritized and broken down by areas of interest.

The meeting adjourned at 4:10 p.m.