February 25, 2002

MEMORANDUM

SUBJECT: Transmittal of the Final Report of the FIFRA Scientific Advisory Panel Meeting Held December 11-12, 2001

TO: Marcia E. Mulkey, Director
Office of Pesticide Programs

William H. Sanders III, Director
Office of Pollution Prevention and Toxics

FROM: Paul I. Lewis, Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Larry C. Dorsey, Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Vanessa T. Vu, Ph.D.
Director
Office of Science Coordination and Policy

Please find attached the final report of the FIFRA Scientific Advisory Panel open meeting held in Arlington, Virginia from December 11-12, 2001. This report addressess a set of scientific issues being considered by the Environmental Protection Agency regarding applicability of the local lymph node assay in dermal sensitization testing and applicability of the up and down procedure methodology for acute oral toxicity testing.

Attachment

cc:

Stephen Johnson
Susan Hazen

Adam Sharp
Janet Andersen
Peter Caulkins
Anne Lindsay
Denise Keehner
Elizabeth Leovey
Linda Moos
Lois Rossi
Frank Sanders
Margaret Stasikowski
Charles Auer
Oscar Hernandez
Antonio Bravo
Douglas Parsons
David Deegan
Don Barnes (SAB)
OPP Docket

FIFRA Scientific Advisory Panel Members
Fumio Matsumura, Ph.D.
Charles Capen, D.V.M.
Herbert Needleman, M.D.
Mary Anna Thrall, D.V.M.

FQPA Science Review Board Members
William Adams, Ph.D.
Mr. Cameron Bowes
Ih Chu, Ph.D.
George Cobb, Ph.D.
George DeGeorge, Ph.D.
James Freeman, Ph.D.
Steven Heeringa , Ph.D.
Michael Luster, Ph.D.
Ernest McConnell, D.V.M.
Peter D.M. Macdonald, D. Phil.
Gary Patterson, Ph.D.

SAP Report No. 2001-13

REPORT

December 11-12, 2001
FIFRA Scientific Advisory Panel Meeting, held at the Sheraton Crystal City Hotel, Arlington, Virginia

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Session 1: Applicability of the Local Lymph Node Assay in Dermal Sensitization Testing

Session 2: Applicability of the Up and Down Procedure Methodology for Acute Oral Toxicity Testing

NOTICE

This report has been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). This report has not been reviewed for approval by the United States Environmental Protection Agency (Agency) and, hence, the contents of this report do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP was established under the provisions of FIFRA, as amended by the Food Quality Protection Act (FQPA) of 1996, to provide advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act Science Review Board members serve the FIFRA SAP on an ad-hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Larry Dorsey, SAP Executive Secretary, via e-mail at dorsey.larry@.epa.gov.

CONTENTS

Session 1: Applicability of the Local Lymph Node Assay in Dermal Sensitization Testing

PARTICIPANTS - 8
PUBLIC COMMENTERS -8
INTRODUCTION - 9
CHARGE -10
PANEL RECOMMENDATIONS - 11
DETAILED RESPONSE TO THE CHARGE - 11
REFERENCES- 14

Session 2: Applicability of the Up and Down Procedure Methodology for Acute Oral Toxicity Testing

PARTICIPANTS - 16
PUBLIC COMMENTERS - 17
INTRODUCTION - 17
CHARGE - 18
PANEL RECOMMENDATIONS - 18
DETAILED RESPONSE TO THE CHARGE - 19
REFERENCES - 26

SAP Report No. 2001-13A

FIFRA Scientific Advisory Panel Meeting,
December 11, 2001, held at the Sheraton Crystal City Hotel,
Arlington, Virginia

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Applicability of the Local Lymph Node Assay in Dermal Sensitization Testing

Mr. Paul Lewis
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: 2/25/02

Fumio Matsumura, Ph.D.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date: 2/25/02

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
December 11, 2001

Applicability of the Local Lymph Node Assay in Dermal Sensitization Testing

PARTICIPANTS

FIFRA SAP Session Chair
Fumio Matsumura, Ph.D., Institute of Toxicology and Environmental Health
University of California at Davis, UC Davis, Davis, CA

Designated Federal Official
Mr. Paul Lewis, FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy

FIFRA Scientific Advisory Panel
Herb Needleman, M.D., University of Pittsburgh, School of Medicine, Pittsburgh, PA
Mary Anna Thrall, D.V.M., Department of Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members
Mr. Cameron Bowes, Health Canada, Pesticide Management Regulatory Agency
Ottawa, Ontario, Canada
Ih Chu, Ph.D., Health Canada, Environmental Health Centre, Ottawa, Ontario, Canada
George DeGeorge, Ph.D., MB Research Laboratories, Spinnerstown, PA
James Freeman, Ph.D. , ExxonMobil Biomedical Sciences, Annandale, NJ
Michael Luster, Ph.D., National Institute for Occupational Safety and Health
Morgantown, WV
Gary Patterson, Ph.D., Department of Pesticide Regulation, California Environmental Protection Agency, Sacramento, CA

PUBLIC COMMENTERS

Oral statements were made by:
Martin L. Stephens, Ph.D. on behalf of The Humane Society of the United States
Abby Jacobs, Ph.D. on behalf of the Food and Drug Administration
Frank Gerberick, Ph.D. on behalf of The Proctor and Gamble Company

Written statements were received as follows:
Abby Jacobs, Ph.D. on behalf of the Food and Drug Administration
People for the Ethical Treatment of Animals

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency issues pertaining to applicability of the local lymph node assay (LLNA) in dermal sensitization testing. Advance notice of the meeting was published in the *Federal Register* on October 5, 2001. The review was conducted in an open Panel meeting held in Arlington, Virginia, on December 11, 2001. The meeting was chaired by Fumio Matsumura, Ph.D. Mr. Paul Lewis served as the Designated Federal Official.

The purpose of this meeting was to seek the SAP's comments on the regulatory applicability of the LLNA, a test method for assessing the allergic contact dermatitis (skin sensitization) potential of chemicals and compounds. The assay was found to be scientifically valid by an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) external peer review as an alternative to traditional guinea pig tests (e.g., Buehler test and Guinea Pig Maximization test). It was also found to provide animal welfare advantages.

In 1996, the SAP approved the incorporation of the LLNA in the Agency's harmonized OPPTS test guidelines (e.g., 870.200 Skin Sensitization) as a screening method. The Agency has now revised its harmonized OPPTS test guidelines to incorporate the LLNA for use as a stand alone method for assessing skin sensitization potential under the appropriate circumstances. These revisions and details of how the LLNA is proposed to fit into both EPA, Office of Pesticide Programs and EPA, Office of Pollution Prevention and Toxics evaluations of skin sensitization were presented to the SAP for comment.

Richard Hill, M.D., (Office of Science Coordination and Policy) and Mr. Jay Ellenberger, Associate Director, Field and External Affairs Division, Office of Pesticide Programs provided welcoming and introductory remarks, respectively. Ms. Debbie McCall (Office of Pesticide Programs) provided an overview of test guideline activities. Karen Hammernik, Ph.D. (Office of Pesticide Programs) summarized the goals and objectives of the local lymph node assay. Jean Meade, Ph.D. (National Institute of Occupational Safety and Health) offered a description of the local lymph node assay procedure. Ms. Debbie McCall (Office of Pesticide Programs) discussed the EPA Office of Pesticide Programs and Office of Pollution Prevention and Toxics position on hypersensitivity testing.

In preparing this report, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. This report addresses the information provided and presented within the structure of the charge by the Agency.

CHARGE

1. The LLNA is generally applicable to testing a wide range of chemicals except for metallic compounds and substances which do not sufficiently adhere to the ear, such as aqueous materials. Compared to currently accepted guinea pig methods, the LLNA:

(a) provides quantitative as well as qualitative data and an assessment of dose-response;
(b) is suitable for testing colored substances since an erythematous response is not produced; and
(c) provides various advantages for animal welfare (e.g. less animal discomfort, shorter test duration, and potentially fewer animals).

Question: Does the Panel agree with the Agency's proposal that, when appropriate, the LLNA should be the preferred dermal sensitization test?

2. ICCVAM recommended and EPA agrees that a concurrent positive control should be included as part of each LLNA to assess proper assay conduct and to use as a standard for comparison of results between studies and laboratories.

Question: Does the Panel agree that a concurrent positive control should be included with each chemical assessed by the LLNA? If not, what does the panel suggest?

3 .EPA proposes that both stimulation index and statistical evaluations of dosed versus control animals be developed to aid in the evaluation of the test outcome.

Question: Has enough guidance been provided in the revised guidelines as to what constitutes a positive hypersensitivity response in the LLNA? If not, what guidance does the Panel suggest?

4. LLNA, guinea pig and human patch test outcomes are available for some 200 chemicals. Results in the LLNA compared favorably with those from the guinea pig. The LLNA was at least as good as the guinea pig test in predicting human responsiveness.

Question: Does the Panel agree that the LLNA guideline can be used for chemicals regulated in both the pesticides and toxic substances programs?

5. The OPPTS guideline builds upon the recommendations made by ICCVAM and it is harmonized with the guideline recently approved in principle by OECD.

Question: Is the draft LLNA test method complete and clearly presented in the revised OPPTS guideline for skin sensitization? If not, please supply guidance for improving the presentation.

PANEL RECOMMENDATIONS

The Panel determined that the LLNA can be used by both the OPP and OPPT, where appropriate. In addition, the Panel agreed with the Agency proposal that the LLNA is applicable for testing chemicals to elicit contact sensitization and should be considered a preferred, stand-alone assay. The Panel noted that the use of LLNA could possibly be expanded to other compounds, such as metals, that were not previously considered appropriate by ICCVAM. The Panel agreed with the Agency that a concurrent positive control should be included with each chemical assayed by the LLNA. In addition, the Panel concluded that a qualitative or quantitative measurement of local irritation should be considered with use of the LLNA. Finally, the Agency should provide a detailed protocol for performing the LLNA based upon the procedure described in the ICCVAM report.

DETAILED RESPONSE TO THE CHARGE

The specific issues to be addressed by the Panel are keyed to the Agency's background document, dated November 1, 2001, and are presented as follows:

1. The LLNA is generally applicable to testing a wide range of chemicals except for metallic compounds and substances which do not sufficiently adhere to the ear, such as aqueous materials. Compared to currently accepted guinea pig methods, the LLNA:

(a) provides quantitative as well as qualitative data and an assessment of dose-response;
(b) is suitable for testing colored substances since an erythematous response is not produced; and
(c) provides various advantages for animal welfare (e.g. less animal discomfort, shorter test duration, and potentially fewer animals).

<u>Question</u>: Does the Panel agree with the Agency's proposal that, when appropriate, the LLNA should be the preferred dermal sensitization test?

The Panel agreed with the Agency's proposal that the LLNA, while representing only the sensitization phase, is applicable to test chemicals for the potential to elicit allergic contact dermatitis and that the LLNA should be considered a preferred, stand-alone assay. This preference is based primarily on the scientific strengths of the assay: (1) it provides a mechanism-based, objective and quantitative endpoint; (2) it evaluates dose response; and (3) and is amenable to statistical analysis. Other attributes include: (1) it has been validated and shown to perform at least as well as guinea pig assays in predicting human response; (2) offers animal welfare advantages, e.g., reduces the number of animals utilized, reduces stress via reduction of the test protocol period and by eliminating the potential distress of a "challenge" response; (3) the assay

provides the ability to test colored substances, and reduces the time/cost of evaluating skin sensitization in laboratory animals.

ICCVAM determined that the LLNA was not appropriate for several classes of chemicals, including metals, strong irritants, and aqueous soluble materials. Data developed since the ICCVAM LLNA report was published suggest that the stated limitations are too restrictive. Thus, recent evidence indicates that use of DMSO or similar polar solvent results in successful identification of 11 of 13 metal salts tested (Basketter, 1999). Strong irritants are difficult to assess in all sensitization assays, but the LLNA may offer some advantages through assessment of dose response and localized irritation at the site of application. Problems with aqueous solutions may be overcome by use of wetting agents. The Panel generally agreed that expanding application of the LLNA to these additional classes of chemicals should be considered. However, one Panel member noted that the SAP could not provide a definitive position since it did not have adequate time to review these new data but agreed that a broader use of the LLNA deserved consideration.

2. ICCVAM recommended and EPA agrees that a concurrent positive control should be included as part of each LLNA to assess proper assay conduct and to use as a standard for comparison of results between studies and laboratories.

Question: Does the Panel agree that a concurrent positive control should be included with each chemical assessed by the LLNA? If not, what does the panel suggest?

The Panel fully agreed that a concurrent positive control should be included with each chemical assayed by the LLNA. As a new method, it is important to include a positive control such as 2- mercaptobenzothiazole or hexyl cinnamic aldehyde in each study for the following reasons: (1) demonstrates the technical capability of the laboratory to conduct such a test; (2) provides inter-laboratory comparison of the LLNA test method; and (3) ability to compare reproducibility between studies within the laboratory.

Some Panel members indicated that at some point in time, conduct of the positive control might be reduced to an interval such as every 6 months. However, it was agreed it would be premature to include this option in the proposed test guideline (e.g. that prior to gaining broad experience and confidence in the LLNA). This question could be revisited by the SAP in a few years when the database is expected to be more robust. In the interim, the Agency should suggest to OECD that its draft guideline be revised to reflect this technical input.

The Panel also considered whether guidance for assay acceptance could be provided by identifying appropriate ranges/criteria for positive control responses. The general view was that this was reasonable conceptually, but no specific proposals were made that gained SAP support.

3. EPA proposes that both stimulation index and statistical evaluations of dosed versus control animals be developed to aid in the evaluation of the test outcome.

Question: Has enough guidance been provided in the revised guidelines as to what constitutes a positive hypersensitivity response in the LLNA? If not, what guidance does the Panel suggest?

Significant discussion focused on the guidance regarding what constitutes a positive response. As worded in the draft guideline, the guidance was interpreted by some of the Panel that a statistically significant response, (e.g. dose response and a statistically significant SI value more than 3) could constitute a "positive." However, this was clarified as not the intent of the Agency and also inconsistent with the interpretation of the validation data by ICCVAM. The principal criterion for identification of a positive sensitizer is an SI greater than or equal to 3. Dose response, statistical evaluations, and assessments of overt toxicity or primary irritation supplement that criterion to provide an overall interpretation of the data. This was judged to be adequate and appropriate by the majority of the Panel. The Panel acknowledged that there are some areas of borderline responses for which interpretation will be difficult, and flexibility rather than a rigid guideline will be required.

The inclusion of a qualitative or quantitative determination of local irritation was deemed critical for several reasons. This determination could include measurement or observation of ear swelling, erythema, etc. Testing should occur at doses that span the range of non-irritating to moderately irritating because some irritation may be a necessary prelude to sensitization, but this irritation needs to be limited. Evaluation of irritation may also help distinguish between irritant and sensitization responses. Further guidance should be provided on interpretation of results when irritation is present, especially if the sensitization response is weak. Also, excessive inflammation may inhibit response, yielding an inverted 'U' shaped dose response curve. One Panel member noted that no guidance is provided on calculation of the required "associated error term" for the SI.

An issue that was unresolved was whether vehicle selection may bias data outcome, and whether guidance for vehicle selection may be needed to achieve the maximum concentration/skin exposure of the test substance. The Panel agreed that the properties of the vehicle can alter the test chemical's solubility, skin penetration, etc. This can result in a altered relative potency.

4. LLNA, guinea pig and human patch test outcomes are available for some 200 chemicals. Results in the LLNA compared favorably with those from the guinea pig. The LLNA was at least as good as the guinea pig test in predicting human responsiveness.

<u>Question</u>: Does the Panel agree that the LLNA guideline can be used for chemicals regulated in both the pesticides and toxic substances programs?

The Panel determined that the LLNA can be used by both the pesticide and toxic substances programs when it is appropriate. The Panel's conclusion is based on the following factors:

(a) LLNA has been validated and shown to perform as well as the traditional Guinea pig assays for prediction of human sensitization potential for a broad range of chemicals including pesticides, industrial chemicals and a limited number of pharmaceuticals;
(b) Guinea Pig Maximization and the Buehler assay are currently considered acceptable under both programs;
(c) The LLNA offers advantages over traditional assays (e.g., colored chemicals/formulations can be tested with the LLNA) and provides quantitative data;
(d) Animal welfare considerations support use of this assay as broadly as possible.

5. The OPPTS guideline builds upon the recommendations made by ICCVAM and it is harmonized with the guideline recently approved in principle by OECD.

Question: Is the draft LLNA test method complete and clearly presented in the revised OPPTS guideline for skin sensitization? If not, please supply guidance for improving the presentation.

The Panel agreed that the draft guideline for the LLNA is overall clearly presented. The Agency should provide a detailed protocol for performing the LLNA based upon the procedure described in the ICCVAM report. Laboratories conducting the LLNA should have minor flexibilities in performing the assay (e.g., means of precipitating DNA or possible use of IP injections in lieu of IV injection of isotopes). However, in these cases, the laboratory must provide adequate justification, preferably in advance, to the Agency. All major protocol requirements must be followed (e.g., strain of mouse, response quantification via scintillation counting) to ensure the adequate conduct of the method as validated. The potential for development of alternative endpoints and other refinements to the method is recognized, but these require validation.

REFERENCES

Basketter, DA. 1999. Identification of metal allergens in the local lymph node assay. American Journal of Contact Dermatitis. 10(4):207-12.

SAP Report No. 2001-13B

FIFRA Scientific Advisory Panel Meeting,
December 12, 2001, held at the Sheraton Crystal City Hotel, Arlington, Virginia

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Applicability of the Up and Down Procedure Methodology for Acute Oral Toxicity Testing

Mr. Paul Lewis
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: 2/25/02

Fumio Matsumura, Ph.D.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date: 2/25/02

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
December 12, 2001

Applicability of the Up and Down Procedure Methodology for Acute Oral Toxicity Testing

PARTICIPANTS

FIFRA SAP Session Chair
Fumio Matsumura, Ph.D., Institute of Toxicology and Environmental Health, University of California at Davis, UC Davis, Davis, CA

Designated Federal Official
Mr. Paul Lewis, FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy

FIFRA Scientific Advisory Panel Members
Charles C. Capen, D.V.M. , Department of Veterinary Biosciences, The Ohio State University, Columbus, OH
Herbert Needleman, M.D., School of Medicine, University of Pittsburgh, Pittsburgh, PA
Mary Anna Thrall, D.V.M., College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members

William Adams, Ph.D., Kennecott Utah Copper Corporation, Magna, UT
Mr. Cameron Bowes, Health Canada, Pesticide Management Regulatory Agency Ottawa, Ontario, Canada
Ih Chu, Ph.D., Health Canada, Environmental Health Centre, Ottawa, Ontario, Canada
George Cobb, Ph.D. The Institute of Environmental and Human Health, Texas Tech University , Texas Tech University Health Science Center, Lubbock, TX
George DeGeorge, Ph.D., M. Research Laboratories, Spinnerstown, PA
James Freeman, Ph.D., ExxonMobil Biomedical Sciences, Annandale, NJ
Steven Heeringa, Ph.D., Institute for Social Research, University of Michigan, Ann Arbor, MI
Ernest McConnell, D.V.M., TOXPATH Inc., Raleigh, NC
Peter D.M. MacDonald, D.Phil., Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
Gary Patterson, Ph.D., Department of Pesticide Regulation, California Environmental Protection Agency, Sacramento, CA

PUBLIC COMMENTERS

Oral statements were made by:
Martin L. Stephens, Ph.D. on behalf of The Humane Society of the United States

Written statements were received as follows:
People for the Ethical Treatment of Animals

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency issues pertaining to applicability of the up and down procedure methodology for acute oral toxicity testing. Advance notice of the meeting was published in the *Federal Register* on November 15, 2001. The review was conducted in an open Panel meeting held in Arlington, Virginia, on December 12, 2001. The meeting was chaired by Fumio Matsumura, Ph.D. Mr. Paul Lewis served as the Designated Federal Official.

The purpose of this meeting is to seek comments of the FIFRA SAP on the regulatory applicability of the Up and Down Procedure (UDP) for acute oral toxicity testing. Acute oral toxicity testing constitutes the adverse health effects that occur within a short time of administration of a single dose of a chemical and provides information on its potential health and environmental hazards and risks. Acute oral toxicity is a basic requirement for registration and reregistration of pesticide active ingredients and products. An improved version of the UDP has been developed as an alternative method for use by member nations of the Organization for Economic Cooperation and Development to meet regulatory needs for acute toxicity. Accordingly, this method will replace the traditional acute oral toxicity test in EPA, Office of Prevention, Pesticides and Toxic Substances (OPPTS) test guideline 870.1100. The test procedure in this guideline is of value in minimizing the number of animals required to determine the acute oral
toxicity testing of a chemical. In addition to the estimation of LD50 and confidence intervals, the test allows the observation of signs of toxicity. Moreover, use of guidance for humane endpoints should reduce the overall suffering of animals in this type
of test. The UDP is to be used for acute oral toxicity testing.

Richard Hill, M.D., (Office of Science Coordination and Policy) and Mr. Jay Ellenberger, Associate Director, Field and External Affairs Division, Office of Pesticide Programs provided welcoming and introductory remarks on behalf of the EPA, Office of Prevention, Pesticides and Toxic Substances management. Amy Rispin, Ph.D., (Office of Pesticide Programs) provided an overview of test guideline activities and development/design of the up and down procedure. Ms. Debbie McCall (Office of Pesticide Programs) presented the performance and applicability of the up and down procedure.

In preparing this report, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. This report addresses the information provided and presented within the structure of the charge by the Agency.

CHARGE

1. Does the SAP agree that the Up-and-Down Procedure (UDP) test guideline generates usable point estimates of the LD50 and confidence intervals that aid in interpreting the significance of LD50 estimates for hazard classification and risk assessment purposes? If not, how might materials be modified to make the method better?

2. Is the conduct of the UDP practical for laboratories to perform in fulfillment of EPA uses? Are there suggestions for improvement?

3. Is the revised UDP method applicable to the regulatory uses for hazard classification for human health and the environment and certain hazard and risk assessment applications under FIFRA? Under TSCA?

4. The OPPTS Harmonized Test Guideline 870.1100 for Acute Oral Toxicity is intended to be used with the AOT425StatPgm software, accompanied by information included in the document titled *Additional Guidance, Toxicology Summary: Performance of the Up-and-Down Procedure*. This guidance describes the strengths and limitations of the UDP regarding estimation of LD50 and confidence intervals. In addition, a software manual for use with the AOT425StatPgm has been provided. Does the 870.1100 guideline and the accompanying manual and other guidance provide sufficient information for study performers and data reviewers? If not, make suggestions for improvement.

PANEL RECOMMENDATIONS

The Panel expects the UDP to give results comparable to the standard LD50 test (OECD 401), when performed under comparable conditions. In addition, LD50 estimates derived from the UDP will be sufficient for hazard classification and labeling of products. The UDP will not be sufficient for risk assessment procedures that require information on the slope or shape of the dose-response curve. This is especially true for tiered ecological risk assessments. Thus, there are limits with application of the UDP for ecological risk assessments. While the Panel concluded that the UDP would be advantageous for hazard assessment and product labeling under the scenarios described, the Panel recommended that the Agency continue to re-evaluate the applications of the UDP after sufficient data have been submitted to determine the adequacy of the methods for purposes of the Agency.

The confidence interval derived from the profile likelihood must be regarded as approximate, especially in cases where the number of animals is very small or where no partial kills were observed, as it may give too high a value for the 95% lower confidence bound for LD50. The same method of calculating the LD50 estimate and confidence interval can be used if additional animals are added to the study design for any reason. In cases where the confidence bounds on a probabilistic risk assessment are too wide, the Agency will have to add additional animals to the UDP or perform a LD50 test with standard techniques.

The Agency must address laboratory and animal management issues of the UDP. These include prolonged study time with few animals dosed at each time, inefficient use of laboratory facilities, maintenance of animal weights and ages over the study; assurance of accurate dosing and dose dilution over the study.

The Guidelines for AOT425StatPgm software should have the following topics moved from "Additional Guidance" to an Appendix: the linear probit dose response model; the likelihood function; the assumption of constant variance; the maximum likelihood estimation of LD50; the stopping rule; and the profile likelihood confidence bounds. In addition, the Guidelines for AOT425StatPgm software should include a schematic outlining the user interface. Finally, the installation procedure for AOT425StatPgm should be simplified to avoid errors and a test data set and protocol be provided to verify correct installation and operation, to be run at installation and periodically thereafter.

DETAILED RESPONSE TO THE CHARGE

The specific issues to be addressed by the Panel are keyed to the Agency's background document, dated November 20, 2001, and are presented as follows:

1. Does the SAP agree that the Up-and-Down Procedure (UDP) test guideline generates usable point estimates of the LD50 and confidence intervals that aid in interpreting the significance of LD50 estimates for hazard classification and risk assessment purposes? If not, how might materials be modified to make the method better?

Some Panelists noted that unlike chemical and physical constants, which are universal, the LD50 should not be regarded as an accurate number. The LD50 for a given test substance could vary many fold depending on animal species, test conditions, etc. It is appropriate to consider the reproducibility of LD50 estimates within the same laboratory, but not more generally than that.

The Panel nevertheless agreed with the Agency that LD50 estimates generated from the Up-and-Down Procedure (UDP) will be sufficient for hazard classification and labeling of products. The point estimates will be especially useful in the classification of mixtures. The confidence interval indicates a range likely to include the true LD50, given the LD50 estimate and its sampling variability. However, due to the small number of animals used, the confidence

interval must be considered inexact and, in many instances, it will be very wide. The Panel nevertheless expects the UDP to give an LD50 estimate comparable to that given by the standard LD50 test per OECD guideline 401, provided both procedures were done under the same conditions.

The simulation results presented to the Panel showed that nominally 95% confidence intervals were generally close to 90% or 95% confidence but in some extreme cases could have actual confidence levels as low as 70%, giving unrealistically high lower bounds.

Even for hazard classification, there is the possibility that a small error in estimation could put a chemical into a different hazard class. The Agency suggested that whenever there was uncertainty at a boundary, the chemical would be assigned to the more toxic class.

For other Agency risk assessment purposes, an understanding of the mechanism of toxicity, the toxicity range and the slope of the dose-response curve is required. While LD50 estimates are good enough for assessing acute human health risk, an estimate of the slope will generally be needed for ecological risk assessment. This limitation is discussed in more detail in the response to Question 3.

The Agency noted that they had examined simulations of the standard LD50 3-dose test for slope and had concluded that it was not reliable, that the slope estimate will be very inaccurate if the "wing" doses are not exactly right. Again, it appears that the UDP may be no less predictive than the standard LD50 test.

Several members of the Panel wanted the UDP protocol to be extended to allow for intermediate doses in cases where only 0% and 100% mortality is observed, with no partial kills, perhaps trying a dose interval of 1.6X. There was, however, a concern that reducing the step size might result in too narrow a range of doses and give a poor LD50 estimate. One Panelist commented that these issues are addressed in the ICCVAM documentation. Furthermore, even when no animals die, it is possible to observe individuals for signs of toxicity and this information can be used in the analysis.

Some Panelists noted that with the UDP, compared to the standard LD50 test, there is less flexibility to adapt the procedure to take account of additional information you may have about the chemicals or the animals, especially in those cases where the UDP uses very few animals. One Panelist suggested that any additional information could be incorporated in a Bayesian statistical analysis, as an alternative to the profile likelihood analysis proposed. A Bayesian analysis, with prior probabilities for slope and LD50 based on experience with similar chemicals and similar animals, might be preferable to using the limiting profile likelihood (infinite slope, or zero sigma) in cases when only 0% and 100% mortality is observed.

2. Is the conduct of the UDP practical for laboratories to perform in fulfillment of EPA uses? Are there suggestions for improvement?

The Panel interpreted the question as asking whether the UDP can be done fairly and readily performed readily, rather than whether the assay is practical. The Panel concluded that the assay can indeed be fairly and readily conducted, inasmuch as any laboratory that can conduct a traditional OECD 401 can conduct the UDP. However, there are some laboratory and animal management issues. These include:

(1) The prolonged study time, with few animals dosed at any given time, utilizes laboratory space inefficiently and adds costs.

(2) The maintenance of animal weights and age over the study time is more difficult. There may be the need to bring in additional shipments of animals. It is anticipated that additional animal shipments would be required infrequently, and the risk to the study would be minimal. This is because suppliers maintain animal health to very high standards. Thus, the risk of introducing disease into an ongoing study is low. Even if such consequences were to occur, the financial cost and consequence of an aborted study is very small. In addition, considering the inherent variability of LD50 estimates, there is also a low probability of introducing significant new error with a new animal shipment.

(3) With respect to special methods that may be needed to ensure accurate dosing and dose dilution over time, several options were discussed. Performing analytical procedures to assess chemical stability, etc. is feasible but adds significant cost (possibly doubling or tripling the cost of the study). In addition, such procedures are not GLP. Preparing fresh dosing solutions and verifying test substance concentrations each day of testing could be problematic for small quantities of hard-to-synthesize research samples.

Specific comments on the protocol were as follows. Selection of the female sex as the default single sex is likely the correct decision if a single sex is to be selected. A dosing interval of 48 hours is appropriate. An age range of 8-12 weeks, while giving needed flexibility for the staggered schedule of this study, may introduce variability because 8 week old rats are still in their rapid growth phase and may be different in dose response than ones of 12 weeks. Therefore, the Panel recommends an older animal at an age of 9-11 weeks. All test animals should be held for a full 14 days for clinical observations and subject to necropsy. However, the suggestion that histopathology be considered in study design was viewed as adding no useful information. More useful pathology will be obtained from studies of longer duration, e.g. 28- and 90-day studies. Selection of starting doses should not be a problem for informed toxicologists.

The Agency does not recommend that additional animals be added to the study design, for example in an attempt to tighten the confidence limits or improve the LD50 estimate. If additional doses are added using an intermediate dose progression scheme, the AOT425 Software Package will calculate the LD50 but would yield a response that the protocol was not followed. Selection of dose progression is the responsibility of the conducting laboratory or sponsor. The profile likelihood calculation of the confidence interval for LD50 will remain valid because it is

conditional on the doses used and does not depend on the protocol used to choose the sequence of doses.

3. Is the revised UDP method applicable to the regulatory uses for hazard classification for human health and the environment and certain hazard and risk assessment applications under FIFRA? Under TSCA?

The revised UDP method test guideline describes a procedure for obtaining an LD50 point estimate and confidence interval that is sufficiently robust for many regulatory uses under both FIFRA and TSCA. These regulatory uses include hazard assessment, classification, and labeling.

The revised UDP method would appear, based on the limited testing the Panel has seen, to provide human health hazard assessment data essentially equivalent to, or slightly improved, over data provided by the traditional LD50 protocol, or the fixed dose and the acute toxic class methods.

However, the revised UDP method has some limitations which limit its applicability within any probabilistic risk assessment, including ecological risk assessment.

(1) The test results do not lend themselves to the generation of a NOAEL or an estimate of the NOAEL by means of a point estimate such as the LD20;

(2) The proposed test does not provide information about the slope of the dose-response curve, especially when a minimal number of animals are used and no partial kills are obtained;

(3) Reduction in the number of animals, while beneficial from one perspective, makes it more likely that the confidence interval will be larger. This raises issues of what is an acceptable confidence interval and when should additional animals be tested.

The consequence of these limitations is that standard LD50 values obtained via this methodology will often have limited utility for risk assessment procedures where exposure and effects curves are compared since a reliable dose-response curve will not be available from the acute rat test without additional testing. Typically, for risk assessments, effects curves generated by probit, logit, or Weibull dose-response analysis are used. The Panel identified this as a limitation of this method and it should be clearly recognized.

Current Agency policy requires a tiered approach for product evaluation within FIFRA. This program is designed to protect wildlife from adverse effects that may be manifested following pesticide exposure. For ecological assessments which rely upon rat studies to estimate risk to other mammals, the Agency must be cautious that free-ranging mammals are not placed at substantial risk. This will happen if the Agency adopts a test procedure that reduces the number of animals used, consequently increasing the uncertainty in the LD50 value. Current ecological

risk assessment procedures are thought to provide some measure of protection for wildlife. At the first tier of the risk assessment process, an LD50 is required for deterministic assessment. This estimate must be accurate enough to prevent erroneous conclusions progressing to higher tiers. Substantial concern exists that uncertainties at the screening level are already too great to provide useful decisions. Known biases in the UDP may impact the validity of assumptions made in current early tier/level ecological risk assessments. If tests which provide data for screening level assessments are altered, the alteration must not reduce the certainty in screening or higher-level assessments.

Subsequent tiers within the risk assessment process use exposure and toxicity data to develop probabilistic risk assessment (PRA). As described in the Agency presentations made during the SAP meeting convened on 13-16 March 2001:

Risk = *f*(exposure, toxicity) = $\Phi(\mu - \beta \log(\text{uptake}))$

Where
Φ = normal cumulative density function
μ = intercept of the log dose probit function
β = slope of the log dose probit function

Lack of a dose response curve might imply an inability to perform a PRA. However, the actual implementation of the PRA attempts to identify a toxicity threshold for sensitive species using the extrapolation method of Aldenburg and Slob (1993). This method can be applied to chronic data (NOEC) or acute data LD 50 values. In the case of acute data, multiple LD 50 values for different (mammalian) species are utilized to identify the 5th most sensitive species. Full dose response data for most of these (focal) species are not typically available and are not needed to develop the acute 5th percentile species. Utilization of LD 50 values generated with the up-down procedure are not expected to significantly affect the outcome of this procedure. However, lack of dose-response data from an up-down test for a key focal species or for the 5th percentile species) would prevent accurate PRA for that species, i.e., it would not be possible to accurately assess effects on that species due to levels above or below the LD 50 exposure level.

PRAs involving the aquatic environment also depend strongly on information that will not be provided by a simple UDP. At the screening level, both LD50 and NOECs are used. Also, Figure 2 of "A Probabilistic Model and Process to Assess Acute Lethal Risks to Aquatic Organisms" prepared for the 13-16 March 2001 SAP Meeting, indicated the need for a log probit fit of toxicity data to obtain "Intercept, slope, standard errors." The paucity of data available for this purpose is a critical issue that the Agency is addressing and that other SAPs have advised the Agency to evaluate.

In summary, altering the LD50 determination will likely not effect hazard classification, but adoption of the UDP in other applications can impair the ability to perform PRAs within the FIFRA process. Without slope information, the dose response for a given species will be largely

unknown and the confidence bound placed on modeled probabilistic risk assessments may be unacceptably wide. This will require either (1) repeating LD50 tests with standard techniques, or (2) collecting exposure and effect data for numerous non-target species during large scale field studies under normal-use scenarios. Given the needed slope data, the UDP will *de facto* become the dose-sighting study to determine slope. These factors point to the need for traditional LD50 tests when performing higher tier ecological risk assessments.

4. The OPPTS Harmonized Test Guideline 870.1100 for Acute Oral Toxicity is intended to be used with the AOT425StatPgm software, accompanied by information included in the document titled *Additional Guidance, Toxicology Summary: Performance of the Up-and-Down Procedure*. This guidance describes the strengths and limitations of the UDP regarding estimation of LD50 and confidence intervals. In addition, a software manual for use with the AOT425StatPgm has been provided. Does the 870.1100 guideline and the accompanying manual and other guidance provide sufficient information for study performers and data reviewers? If not, make suggestions for improvement.

The direct response to this question is that the combination of the Guidelines, the Additional Guidance, the software and the software documentation do provide sufficient guidance to enable laboratory specialists to conduct the test. These materials, supplemented with the special simulation study results and two-volume ICCVAM report, provide reviewers of the data with a general basis for evaluating the UDP estimate of the LD50. Except for statistically pathological cases such as linear dose-response curves with low slopes, nonlinear dose-response curves, or irregular profile likelihoods from small samples, the materials also provide the data reviewer with sufficient information to evaluate the approximate confidence bounds for the estimated LD50 and its corresponding application in hazards classifications. Interpretation of these results in the more pathological cases will require expertise and advice that is probably beyond the basic guidance that such documents can be expected to provide.

The quality of the written guidelines and documentation materials is adequate but as is almost always the case, could be improved. The outline of the Guidelines does not lend itself easily to cross-referencing background and procedural information with supplementary technical guidance, but this may be a function of the established formatting rules for EPA guidelines. The Guidelines section on Background, Initial Considerations and Principles is certainly sufficient. In the Procedures section, Table 1 illustrating default dosing levels could be brought forward to accompany the text that describes the specification of doses for sequential UDP steps. Section (7) on observations is sufficient, with the one minor exception, the definition of "time of death" for moribund or suffering animals that are euthanized. Is it time at euthanization or a projection of the likely time to mortality without intervention?

The illustration of the procedural sequence and calculations for three test scenarios provided in Table 2-4 are valuable but since they are modeled on the testing software, care should be taken to update the guidelines to the current version of the AOT425StatPgm software. The Panel recommends that the Agency consider integrating the "Additional Guidance" into a

technical appendix to the Guidelines. A list of statistical and technical topics that could be consolidated in the Appendix includes:

(1) the linear probit dose response model;
(2) the corresponding likelihood function;
(3) the assumption of constant variance in the maximum likelihood estimates (MLEs);
(4) the MLE of LD50;
(5) the stopping rule; and
(6) the computation of confidence bounds through profile likelihood.

Regarding (5) and (6), it may be useful to show the relationship of the default stopping rule to the method for determining the confidence interval bounds by the profile likelihood method.

Note that the CI under the profile likelihood is derived by "pivoting" on a 1 degree of freedom likelihood ratio test comparing the likelihood at the MLE to the likelihood at alternative specified values of the LD50:

-2(CRIT LL - MAX LL)

is approximately distributed as a chi-square on 1 degree of freedom, for which the 95% critical value is 3.84, the confidence interval is defined by

MAX LL - CRIT LL < 3.84/2 = 1.92

where CRIT LL is the log likelihood under the probit model at the lower or upper confidence interval bound values of LD50 and MAX LL is the log likelihood at the MLE estimate of LD50.

The stopping rule is invoked after 4 "reversals" in up-down dosing outcomes, where the number 4 was presumably chosen to set minimum stability and bounding for the MLE of LD50. The criterion for stopping is based on an evaluation of the probit likelihood at two points, one below and one above the current MLE of LD50. The recommended points are the MLE/2.5 and 2.5*MLE. If the ratio of the MLE likelihood to the likelihood evaluated at both of these two test bounds exceeds 2.5, the testing stops. By inspection of the two-sided stopping criteria, the stopping rule can only be invoked when the likelihood for the data is concave about the current MLE of the LD50. Transforming this criterion to a log likelihood scale demonstrates that the choice of a > 2.5 likelihood ratio implies that the stopping criteria is implemented when the 95% profile likelihood confidence interval is well defined.

The stopping rule can be written:

$\{L(LD_{50})/L(LD_{50}/2.5), L(LD_{50})/L(LD_{50}*2.5)\} > \{2.5, 2.5\}$

or, in terms of lnL,

MAX LL – TEST LL > ln(2.5) > .916

for both the lower and upper test values.

The documentation for the AOT425StatPgm software for implementing the UDP dosing and LD50 calculations is generally well written and provides sufficient guidance to the user population. There are several improvements/enhancements that the SAP would like to suggest to the EPA. First, the installation procedure involves user changes in installation file names and updates require users to "ignore" error messages. For a program of this caliber and potential uses, these user manipulations should not be needed; they only serve to increase the probability of installation errors including incorrect linkage to program libraries or data tables. The installation procedure should be simplified to make it efficient and error-free from the user's perspective.

Second, it might be more accurate for the documentation to describe the statistical analysis as "nonlinear" rather than "computationally intensive" as the latter term usually refers to Markov Chain Monte Carlo and resampling methods.

Third, it would be valuable to include an initial schematic that outlines the hierarchical relation of the two main program windows (Data Edit and Report), the task bars associated with each window and menu choices (where applicable) associated with each button on the task bar. This schematic would provide the user a quick overview of the full user interface of the program and would facilitate migration between windows, tasks and options during program use.

Finally, the SAP recommends that the Agency establish a test data set and test protocol that users may apply to verify the correct installation of the software and to verify periodically that the program calculations match the test standard.

LITERATURE CITED

Aldenberg, T. and W. Slob, 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicol. Environ. Saf.* 25:48-63.