SAP Report No. 2000-02
August 2, 2000

REPORT

FIFRA Scientific Advisory Panel Meeting, April 5-7, 2000, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Session I - Implementation Plan for Probabilistic Ecological Assessment: A Consultation

Session II - Insect Repellent Product Performance Testing Guideline Evaluation

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 EPA 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.

TABLE OF CONTENTS

Session I: Implementation Plan for Probabilistic Ecological Assessment: A Consultation

Session II. Insect Repellent Product Performance Testing Guideline Evaluation

SAP Report No. 2000-02
August 2, 2000

REPORT

FIFRA Scientific Advisory Panel Meeting, April 5-7, 2000, held at the Sheraton Crystal City Hotel

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

Session I - Implementation Plan for Probabilistic Ecological Assessment: A Consultation

Session II - Insect Repellent Product Performance Testing Guideline Evaluation

SAP Report No. 2000-02A, August 2, 2000

REPORT:

FIFRA Scientific Advisory Panel Meeting, April 5-6, 2000, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Implementation Plan for Probabilistic Ecological Assessment

Ms. Laura E. Morris
Designated Federal Officials
FIFRA/Scientific Advisory Panel
Date:_____

Ronald J. Kendall, Ph.D.
Chair
FIFRA/Scientific Advisory Panel
Date:_____

Federal Insecticide, Fungicide, and Rodenticide Act

Scientific Advisory Panel Meeting
April 5-6, 2000

Session I: Implementation Plan for Probabilistic Ecological Assessment

PARTICIPANTS

Chair
Ronald J. Kendall, Ph.D., Director and Professor, The Institute of Environmental and Human Health, Texas Tech University/Texas Tech Health Services Center, Lubbock, TX

Permanent Panel Member
Mary Anna Thrall, D.V.M., Professor, Department of Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, Colorado

FQPA Science Review Board Members
William Adams, Ph.D., Assistant Professor, University of Minnesota School of Public Health Minneapolis, MN
George P. Cobb, Ph.D., Texas Tech University and the Texas Tech Health Science Center, Lubbock, TX
Paul W. Eslinger, Ph.D., Staff Scientist, Pacific Northwest National Laboratory, Richland, WA
Kirk Hatfield, Ph.D., University of Florida, Gainesville, FL
Elwood F. Hill, Ph.D., Adjunct Professor, Gardnerville, NV
Thomas W. LaPoint, Ph.D., Professor, Institute of Applied Science, University of North Texas Denton, TX
Dwayne Moore, Ph.D., The Cadmus Group, Ottawa, Ontario Canada
Kenneth M. Portier, Ph.D., Associate Professor, Department of Statistics - Institute of Food and Agricultural Sciences, Gainesville, FL
Thomas Potter, Ph.D., USDA/ARS, University of Georgia, Southeast Watershed Research Laboratory, Tifton, GA
Harold Van Es, Ph.D., Cornell University, Department of Soil, Crop and Atmospheric Sciences Ithaca, NY
Bill Williams, Ph.D., Kennedy-Jenks, Consultants, Portland, OR
Barry Wilson, Ph.D., Department of Environmental Toxicology, University of California-Davis Davis, CA

Designated Federal Official
Ms. Laura E. Morris, FIFRA Scientific Advisory Panel, Office of Prevention, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, DC

PUBLIC COMMENTS

Oral statements were received from:
Paul Hendley, Ph.D., Zeneca Ag Products on behalf of ECOFRAM
Jeffery Giddings, Springborn Laboratories, Inc. on behalf of ECOFRAM
Ray McAllister, on behalf of the American Crop Protection Association

Written statements were received from:
Paul Hendley, Ph.D., Zeneca Ag Products on behalf of ECOFRAM
Jeffery Giddings, Springborn Laboratories, Inc. on behalf of ECOFRAM

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 regarding the review of issues pertaining to the implementation plan for probabilistic ecological assessments. The purpose of this consultation was to provide the SAP with a progress report regarding the Agency's initiative to revise the ecological assessment process. Following the recommendations from the May 1996 SAP meeting and building on previous efforts within the Environmental Fate and Effects Division (EFED) of the Office of Pesticide Programs, EFED began a new initiative in 1997 to revise the ecological assessment process. The main focus of this initiative is to identify, develop , and validate tools and methodologies to conduct probabilistic ecological assessments and improve risk characterization. Advance public notice of the meeting was published in the Federal Register on March 16, 2000. The review was conducted in an open Panel meeting held in Arlington, VA, on April 5-6, 2000. The meeting was chaired by Ronald J. Kendall, Ph.D. Ms. Laura Morris served as the Designated Federal Official.

CHARGE

The specific issues to be addressed by the Panel are keyed to the background documents, ***"Technical Progress Report of the Implementation Plan for Probabilistic Ecological Assessments: Aquatic Systems", and "Technical Progress Report Implementation Plan for Probabilistic Ecological Assessments: Terrestrial Systems"***.

Part I: Questions on Aquatic Technical Progress Report

1. The interim exposure model proposed for Level 1 is GENEEC. This model essentially simulates direct application of a pesticide to a 1 hectare, 2 meter deep pond. GENEEC generally yields higher estimated concentrations than those found in the environment, but occasionally monitoring data does exceed GENEEC estimates. Does the Panel believe that it would be a good use of resources to pursue a more conservative (e.g., shallower water body which may be more representative of small, sensitive areas) scenario or does the GENEEC appear to be conservative enough? Please provide rationale.

2. Do the various tests and range of aquatic species tested in Level 1 and 2 appear to be sufficient to protect more sensitive species within taxa when extrapolation factors are applied, given potential large differences in sensitivity based on mode of action? Please provide guidance.

3. In reference to amphibians,

A. EFED is proposing the addition of amphibian testing when a small margin of safety exists (less than one order of magnitude) between expected concentrations of a chemical and its toxicity in other taxa? Does the SAP agree with this approach and/or are there additional factors which should also be considered.

B. EFED is proposing to use the Frog Embryo Teratogenesis Assay-*Xenopus* (FETAX) test as an interim amphibian test model. Do the Panel members agree and do the Panel members have any additional or alternative suggestions to make in this area?

4. Does the SAP agree with using a regression-based approach to evaluate chronic tests instead of the currently used hypothesis testing approach, where the data support this analysis (regulatory endpoint of EC_X vs NOAEC)? Please provide rationale.

5. ECOFRAM noted that chronic endpoints such as hatching success may be affected by short-term exposures at critical life-stages, while endpoints such as growth may be more reflective of cumulative exposure. Therefore, they suggested that the Level 1 chronic risk quotient be the ratio of the model-estimated peak EEC to the EC_X or NOAEC to reflect the need to be protective at Tier 1. Does the SAP agree with this use of a peak exposure as a chronic effects screening measure in lieu of using a time-weighted average at Level 1? Please provide rationale and guidance.

6. For exposure modeling, EFED's preferred current approach is to use actual historical weather data instead of using a random weather generator? Does the Panel agree with this approach and does the Panel have any additional suggestions or proposals that would improve this?

7. In reference to variability in model input parameters,

A. Is the consideration of variability in PRZM/EXAMS model input parameters through Monte Carlo analysis at Level 2 useful, or should this consideration be taken up at Level 3? Please discuss.
B. Would the SAP suggest any other approaches to address this variability?

8. Regarding the derivation of a common slope factor for extrapolations in Level 2 dose-response estimates to evaluate effects on more sensitive species, does the SAP have a recommendation for an approach for deriving a generic slope?

9. Does the SAP have recommendations on sediment toxicity testing, especially regarding appropriate level of assessment, and benthic fish testing species/protocols? Please be specific.

10. Does the SAP have recommendations regarding population and community models which might be most suitable for regulatory evaluations?

11. Does the SAP agree with considering regional evaluations at Level 3, and focusing on 90% crop/use scenarios at Level 2, in order to direct initial evaluations toward high-end risk sites? Please discuss.

12. In reference to species sensitivity distributions,

A. How many species within a taxa should be tested at Level 3 to adequately characterize a species sensitivity distribution without the need for extrapolation factors?

B. How should the aquatic taxa be grouped for evaluating species sensitivity distributions?

13. In reference to additional sublethal effects testing,

A. Does the SAP concur that additional sublethal effects testing (such as immunocompetence) at Level 3, when a specific mode of action is of concern, will improve risk evaluations? Please provide rationale.

B. Are there specific sublethal effects that the SAP would recommend as most useful to assess? Of particular interest are sublethal effects that may affect endpoints potentially suitable for regulatory decision-making (e.g., survival, fecundity).

Part II: Questions on Terrestrial Technical Progress Report

1. The Terrestrial ECOFRAM Workgroup recommended that the LC_{50} test be modified by calculating an incipient LC_{50}, defined by the point on the study when the LC_{50} does not decrease by more than 1% over two days. (This modification was proposed through OECD as described in background Document #4.) Can the SAP comment on the utility of the recommended LC_{50} test modification to account for exposure durations for pesticides with moderate to long residue dissipation rates in wildlife foods, soil, and drinking water sources?

2. EFED has proposed two approaches for discerning appropriate exposure duration windows for calculating cumulative or time-weighted average exposures for short- and medium-term lethality risk assessments. These are (1) setting the window to match toxicity test duration and (2) extending the window over protracted period limited to some minimal effect point.

A. Can the SAP express a preference for either approach and provide a rationale?

B. If not, can the SAP define a more appropriate approach that utilizes existing data sets (i.e., without resorting to additional data requirements for pharmacological/pharmacokinetics data)?

C. Should exposures (in the absence of additional data) be calculated as averages or cumulative over the assigned exposure window? Please provide the rationale.

3. Because of the present lack of avian inhalation and dermal toxicity data, EFED has proposed an equivalency factor approach, based on laboratory rodent acute toxicity potency comparisons across exposure routes, to normalize exposures from dermal and inhalation routes for birds.

A. Does the SAP believe that, in the absence of specific pharmacological/pharmacokinetics data, that this approach is reasonable and if so, why?

B. If not, can the SAP provide insight into quantitative methods, using the existing data sets, to facilitate a comparison of different exposure routes to the existing toxicity data presently required for birds?

C. There is recognized need, in higher Level of Refinement assessments, for additional toxicity data for routes of exposure other than oral to reduce extrapolation uncertainties. Can the SAP provide any detailed guidance on developing testing protocols (technical points to consider or knowledge of existing methods) for avian effects in birds via the dermal or inhalation routes?

4. Published studies suggest that for some pesticides and birds, exposure via preening can be an important route. Can the SAP comment on the need for considering this route of exposure? If the route is believed to be appropriate for consideration, is the SAP familiar with any quantitative

methods for including this route into the overall exposure assessment?

5. EFED is trying to determine the best basis for acute effects characterization in the risk assessment process. Can the SAP provide guidance on selection of either the acute single oral dose or the dietary toxicity study as the basis for acute lethality assessments?

6. In the opinion of the SAP, what are the minimum study requirements for each Level of Refinement for both lethal and reproductive effects? How many species should be tested at each Level of Refinement to adequately characterize a species sensitivity distribution?

7. EFED has recommended modifications to the avian reproduction test to provide dose-response information for sensitive endpoints. Can the SAP provide suggestions for protocol design and/or the most important considerations that should be factored into the design of the study?

8. EFED has proposed options for interspecies extrapolation factors for both lethal and reproduction risk assessments.

A. Can the SAP comment on whether the proposed method is appropriate?

B. Can the SAP recommend an alternative or additional approach, with supporting detail?

C. Is the proposed approach for reproduction effects appropriately conservative under the existing limitations of available data, and if so, why?

D. If not, does the SAP have a preference for an approach for reproduction effects extrapolations at this time?

9. Can the SAP provide guidance on what additional species would be the most appropriate for testing for both lethal and reproductive effects at the higher Levels of Refinement?

10. The Terrestrial Technical Progress Report presents methods for predicting dose-response slopes for extrapolated sensitive species.

A. Does the SAP have suggestions on ways to improve these methods or can an alternative approach be taken?

B. Can the SAP suggest methods for similar extrapolations for reproduction effects?

11. Can the SAP provide guidance on the methods for developing natural history information on avian species in agro-ecosystems that would be appropriate for use in probabilistic assessments? This would include information on avian census, time budget and dietary proportions on and off treated fields, available food sources, and others.

Closing Question

1. What approach would the SAP recommend to move toward validating the risk assessment processes presented in the Technical Progress Reports? Please provide specific recommendations.

PART I: AQUATIC TECHNICAL PROGRESS REPORT

DETAILED RESPONSE TO THE CHARGE

Questions on Aquatic Technical Progress Report

1. The interim exposure model proposed for Level 1 is GENEEC. This model essentially simulates direct application of a pesticide to a 1 hectare, 2 meter deep pond. GENEEC generally yields higher estimated concentrations than those found in the environment, but occasionally monitoring data does exceed GENEEC estimates. Does the Panel believe that it would be a good use of resources to pursue a more conservative (e.g., shallower water body which may be more representative of small, sensitive areas) scenario or does the GENEEC appear to be conservative enough? Please provide rationale.

Panel members agreed that GENEEC could be improved and that the Agency should invest the necessary resources for improvement. At the same time, several Panel members questioned whether or not continued use of GENEEC at this stage of development was appropriate. The Agency has recently chosen to use monitoring data and watershed scale regression analyses when conducting assessments of potential pesticide concentrations in reservoirs used as drinking water supplies. Another suggestion was that the risk model be reduced to three tiers. In this case, the current Tiers 1 and 2 would be combined and PRZM/EXAMS would be used to assess potential exposures under a variety of pesticide use, geographic and climatic scenarios. The modeling framework is already in place and the effort required to run the simulations is manageable.

Thus, the Panel did not agree that GENEEC is sufficiently conservative to capture exposure dynamics under many pesticide use scenarios. It was noted that runoff and groundwater in many agricultural watersheds is discharged directly to first-order streams. This also may be the case in urban watersheds. Other hydrologic settings where the level of dilution of runoff may be considerably less than is assumed in the model are prairie pothole lakes, playa lakes in the Western USA, and amphibian habitats in general. Amphibians often dwell in wetlands where shallow depressions that accumulate runoff are important habitats. In these areas, dilution of runoff may be minimal; thus, GENEEC would significantly underestimate exposure. There also were concerns that GENEEC may not capture the effects of long-term use and accumulation, particularly for chemicals exhibiting sediment partitioning. In this case, the model will under predict concentrations and associated chronic risks. Also, if the product half life is longer than the application interval, or if significant quantities of toxic metabolites are formed, additional conservatism may be necessary.

One Panel member expressed the opinion that GENEEC has the potential to significantly overestimate exposure in estuarine environments where tidal flushing would provide large dilution factors. For most situations, GENEEC in its current form is appropriately conservative. However, there is concern about relying on a model that uses a single exposure scenario to

represent all pesticide use scenarios. The need for a model that addresses exposure dynamics under a range of hydrologic and pesticide use settings was highlighted.

The following recommendations were provided by the Panel members:

1. There was consensus that GENEEC could be improved by expanding the number of crops, soils, and climates considered. Continued use of a single "cotton scenario" was considered unacceptable.

2. A need for model verification was identified.

3. One method to answer the question about conservatism of the model is to use existing registrant data to determine whether compounds that pose significant ecological risks are indeed identified by the model/protocol.

2. Do the various tests and range of aquatic species tested in Level 1 and 2 appear to be sufficient to protect more sensitive species within taxa when extrapolation factors are applied, given potential large differences in sensitivity based on mode of action? Please provide guidance.

The Panel generally agreed that the model is protective of untested sensitive species. Several points in support of this position are presented below.

Levels 1 and 2 require a total of 9 acute tests (freshwater and marine combined) plus a possible three more for benthic testing. In addition, four chronic tests are required plus a screening level reproduction study. This exceeds EPA's Office of Water requirements for establishing a water quality criterion, an approach that is widely recognized as providing a defensible number for the protection of aquatic life (both across and within taxa).

The available literature suggests that having tested a fish, a *Daphnia* and an alga, one would expect to be within an order of magnitude or less of estimating the most sensitive species on an acute basis (Kimerle et al. 1985). Within a given taxa one would expect to have an even better estimate of what the acute value would be for the most sensitive species, assuming test data across all taxa are used. In Levels 1 and 2, fish and invertebrate categories have two or more required tests. This use of multiple species further increases the potential that sensitive species would be protected by the derived value.

The data of Mayer and Ellersieck (1992) indicates that the use of a single sensitive species (*Daphnia*) reduces the variability to a factor of 15 between the lowest and highest LC(EC) 50 value for 95% of the chemicals tested. Additionally, their data also show that a testing approach (like that proposed by EFED in Tiers 1 and 2) that uses three sensitive species, such as rainbow trout (*Daphnia magna)*, and an amphipod will result in one of the three species being the most sensitive 88-90 % of the time. If one were to assume that this is acceptable, no extrapolation

factor would be necessary. Recognizing that EFED uses more than three test species in its risk assessment, the probability of including the most sensitive species in the data set appears to be even greater than 90%. EFED could sequentially evaluate the potential to obtain an EC or LC 50 value equivalent to the most sensitive value in given data sets 4, 5, or 6 as (etc.) species are used in various Tiers. It is recommended that EFED update the data set of Mayer and Ellersieck (1992) with its available pesticide data, to determine the need for additional extrapolation factors or testing.

A key remaining question is whether or not the standard data set is sufficient to protect threatened and endangered (T&E) fish species. Brix et al. (1999) assessed the sensitivity of 12 T&E species as compared to standard test species (fathead minnow, rainbow trout and sheepshead minnow) for five substances (carbaryl, copper, nonylphenol, pentachlorophenol, and permethrin). Additionally, they compared the most sensitive acute toxicity value in the T&E data set with final acute value derived using EPA water quality criteria methodology. The conclusion from this review is that standard fish species do a good job estimating the toxicity of T&E species if rainbow trout are included in the data set. There were a couple of exceptions to this, in which case a safety factor of 2 applied to the rainbow trout acute value was protective of T&E species. Additionally, in all cases where invertebrates were included in the data set analysis, the final acute value was protective of all T&E fish species.

Solomon and Giddings (2000) estimated the probability of detecting a species that is more sensitive than a given sensitive species in a data set of 63 acute toxicity tests with diazinon. The probability of untested species X being more sensitive than the lowest invertebrate species (*Ceriodaphnia dubia*) is 3% and <0.1% for the most sensitive fish species (Novartis Crop Protection 1997).

Several microcosm studies have indicated that effects are rarely seen below the 25th quartile of species tested in the laboratory and present in the test. However, the Panel expressed concern that the Agency did not specify the size of the extrapolation factor, or how it would be used, and had not carefully considered whether or not there was a need for extrapolation factors. Several shortcomings were identified.

Tests need to take into account the nature of the chemical, the mode of toxic action, and the distribution of species relative to potential patterns of exposure. Given the use patterns of a chemical, some test species may never be exposed.

Extrapolation factors should be determined empirically and there should be some flexibility in the selection of test organisms depending on chemical properties such as K_{oc}. A need for tests on benthic species was identified when a pesticide is likely to partition strongly to sediments (high K_{oc}).

Tests are needed for invertebrate species based on the duration of their life-cycle. A 4-day exposure will be a much greater period in the life history of cladoceran invertebrates (short-lived) versus fish or larger invertebrates.

Specific requirements for testing an invertebrate species that reproduces only once annually and increasing the number of plants tested for toxic effects of herbicides were identified.

3. In reference to amphibians

A. EFED is proposing the addition of amphibian testing when a small margin of safety exists (less than one order of magnitude) between expected concentrations of a chemical and its toxicity in other taxa. Does the SAP agree with this approach and/or are there additional factors which should also be considered?

The Panel was supportive of EFED's amphibian testing approach noting that it is valid for the same reasons that testing other species are potentially affected. Pesticides are often applied during the period of egg laying and tadpole development. If the pesticide is being applied or will be applied in areas and at times when amphibians are in early life stages (e.g., hatching, tadpole growth and metamorphosis), then toxicity tests with amphibians should be required whatever the margin of safety shown with other required tests. Another suggestion was that tadpoles should be examined for the possibility of accumulating pesticide residues. They are an important food source for predatory vertebrates thus direct dietary poisoning or biomagnification are concerns.

As an additional reference, the Panel suggested the Agency consider information in the publication *Ecotoxicity of Chemicals to Amphibians: Volume I, Handbook of Ecotoxicity Data*; Devillers and Exbrayat, Gordon and Breach Science Publishers.

B. EFED is proposing to use the Frog Embryo Teratogenesis Assay-*Xenopus* (FETAX) test as an interim amphibian test model. Do the Panel members agree and do the Panel members have any additional or alternative suggestions to make in this area?

The Panel agreed that in the absence of protocols on other amphibian endpoints (e.g., reproduction, growth, survival), the FETAX test is a reasonable interim test model. Because teratogenesis is not as ecologically relevant an endpoint as survival, growth, and reproduction, effort should be directed at developing standardized test protocols for amphibians with these endpoints. There was also agreement that under some circumstances *Rana sp.* can be used effectively with testing limited to immature life stages. Experience has shown that the adults are often much more tolerant of oral pesticide doses than other organisms.

4. Does the SAP agree with using a regression-based approach to evaluate chronic tests instead of the currently used hypothesis testing approach, where the data support this analysis (regulatory endpoint of EC_X vs NOAEC)? Please provide rationale.

There was strong support for a regression-based approach because test statistics can be used to determine whether model fit is adequate and whether the assumptions of the analysis have been met. The utility of the approach in a probabilistic approach to pesticide risk assessment was also emphasized.

As a treatment regime, an exposure-response (EC_x) experimental design was recommended. When using this model at least three, preferably five, concentrations should be used, with at least two replicates per concentration. The selected concentrations should be based on expected effects and not on the EEC, although the selected concentrations are likely to bracket the EEC. This design offers a wider use of these data, useful in extrapolating the results to other concentration levels or to different systems.

One recommended method of multi-variate analysis is Principle Response Curves. It has been proposed for use by several European nations. This analysis takes into account populations of different species, and variances between replicates, between time-points and between treatments. This technique is able to detect direct and indirect effects as well as population recovery.

Several Panel members reported that OECD and others groups are currently revising bioassay protocols to be more amenable to regression analysis but that there are still some issues to be worked out such as choice of model (Bailer and Oris, 1997), spacing of treatments, and incorporation of time to effect information.

5. ECOFRAM noted that chronic endpoints such as hatching success may be affected by short-term exposures at critical life-stages, while endpoints such as growth may be more reflective of cumulative exposure. Therefore, they suggested that the Level 1 chronic risk quotient be the ratio of the model-estimated peak EEC to the EC_x or NOAEC to reflect the need to be protective at Tier 1. Does the SAP agree with this use of a peak exposure as a chronic effects screening measure in lieu of using a time-weighted average at Level 1? Please provide rationale and guidance.

Several Panel members noted that this is a conservative approach and endorsed its use in the first levels of the risk assessment framework. To use concentrations other than peak values may allow pesticides with suggested risks to go untested. Using peak exposures in Level 1 for estimating chronic risks is acceptable, provided the Agency realizes that many pesticides will likely screen through to Level 2.

The dynamics of exposure were identified as an important consideration in support of using peak concentrations. Toxic responses that would be observed if an organism is exposed to a peak concentration followed by much lower levels may not be discernable if only the time-weighted average is considered. Another argument in favor of using peak concentrations was that it may account for toxic degradation products not being considered in the exposure model. The Agency was directed to a recent presentation by Dr. Hendley, an ECOFRAM member. His recommendation would move the exposure assessment to PRZM/EXAMS (i.e. Level 2) if the half life exceeded 0.5 yrs, K_d exceeded 2, and application intervals were less than 4 weeks.

One Panel member recommended the Agency consider the use of a reasonable maximum exposure value (RME) as opposed to the maximum (peak) exposure value. This avoids issues associated with extreme values that may drive the assessment and may in fact be outliers. The

RME approach is now used in exposure assessments in the EPA superfund program.

6. For exposure modeling, EFED's preferred current approach is to use actual historical weather data instead of using a random weather generator? Does the Panel agree with this approach and does the Panel have any additional suggestions or proposals that would improve this?

There was consensus that the use of actual weather data is a superior approach and that increasing the length of the historical record from 35 to 51 years would enhance the value of model outputs. The rationale was that the shorter the historical record, the less likely models will capture temporal correlations which may occur at long return intervals. Monte Carlo type analysis, such as shuffling years, was recommended as a way to try to produce sequences of weather events that are less likely to occur than those obtained in the historical record. A concern was expressed that weather data based on daily record does not allow simulation of runoff from intense short duration (1 hourly) storm events. Such storms contribute disproportionally large amount of pesticides in runoff. Hourly weather data should be pursued that allows for such simulations. If a weather generator can provide event-based specificity, then it will be superior.

7. In reference to variability in model input parameters,

A. Is the consideration of variability in PRZM/EXAMS model input parameters through Monte Carlo analysis at Level 2 useful, or should this consideration be taken up at Level 3? Please discuss.

Panel members responded by stating that it was their understanding that Level 2 analyses were designed to be probabilistic and to be predictive while still using some generic input parameters and stylized species exposure considerations. Thus, they concluded that it is appropriate to use Monte Carlo analysis at Level 2. This position was qualified with cautionary notes and recommendations:

1. In order to make this exercise feasible, generic high exposure scenarios need to be defined (e.g., static headwaters for different crops and regions). This would facilitate the development of standardized distributions for many model input variables (e.g., climate and receiving environment variables), thus reducing analysis time and improving consistency between assessments.

2. If the Agency combines Levels 1 and 2, as a first step, Monte Carlo analysis of the variability of the PRZM/EXAMS model should be postponed to the next level.

3. The Agency should be clear about the need to specify whether the Monte Carlo analysis is being performed to explore the effect of parameter imprecision on risk or the effect of parameter (spatial/temporal) variability on risk. It makes a difference in the types of distributions used and the interpretation of the results. In some cases, the Monte Carlo

is performed on a combined variability term which confounds the two types of variability and makes interpretation more difficult.

4. In the case of PRZM/EXAMS, Monte Carlo analysis is being performed on a multi-parameter function. The Agency should ensure that the acceptable region for combinations of parameters does not contain combinations that would lead to unreasonable scenarios. It is not enough to specify upper and lower bound values for each parameter and then use the parameter space as the simple product of the individual parameter ranges.

B. Would the SAP suggest any other approaches to address this variability?

Techniques such as first order error analysis could be considered with simpler models; however, Monte Carlo analysis is likely the most feasible tool for propagating uncertainties in complex multi-parameter models like PRZM/EXAMS. There are other methods, such as closed-form theoretical methods or analytical propagation techniques, for conducting uncertainty analyses, but these quickly become intractable with complex models.

If the goal is to explore the consequences of uncertainty about choice of distribution, distribution parameters and/or dependencies, tools such as 2nd order Monte Carlo or Probability Bounds Analysis should be considered. These tools can be used to develop bounds on exposure or risk curves – the tighter the bounds the less uncertainty.

The MUSCRAT model was proposed by ECOFRAM as a tool for expanding variability in fate and transport between regions. Because MUSCRAT is a multiple scenario tool, it would seem appropriate to validate and use the model as rapidly as possible.

8. Regarding the derivation of a common slope factor for extrapolations in Level 2 dose-response estimates to evaluate effects on more sensitive species, does the SAP have a recommendation for an approach for deriving a generic slope?

Panel members felt that it was difficult to answer the question because there is a wide range of slope factors for different species exposed to the same chemical. Further, there are insufficient data available to determine whether a geometric or arithmetic mean is most suitable. In the absence of more data, use of the geometric mean was identified as the better approach. However, if slopes differ considerably between test species, EFED may want to consider a distributional approach for estimating concentration-response relationship for generic or 5th percentile species. This would involve having a distribution for LC_{50s}, a distribution for slopes, and an estimate of the correlation between the two parameters.

9. Does the SAP have recommendations on sediment toxicity testing, especially regarding appropriate level of assessment, and benthic fish testing species/protocols? Please be specific.

The extent to which a compound will partition from the aqueous phase to the sediment was identified as the key consideration in determining the need for testing benthic species. There was a consensus that compounds with high K_{oc} or K_{ow} should be screened for benthic fish or invertebrate testing at Level 1. A recommendation was made that sediment tests be conducted with a benthic organism that lives in the sediment and ingests sediments when a substance with a large K_{oc} (sediment-water partition coefficient; i.e., $>10^4$) is being evaluated.

An alternative approach is to use an equilibrium partitioning approach (for non-polar organics and metals). It allows prediction of the sediment pore water concentrations. Therefore, the potential for toxicity to occur can be assessed by calculating the concentration of the pesticide in the pore water and comparing this value to the most sensitive acute or chronic aquatic toxicity test result available for the test substance. This approach assumes similar sensitivities between pelagic and benthic biota and it should work well for epi-benthic biota. It may not work well for biota that ingest sediment.

Selection of relevant treatment levels (for one application) is a problem if multiple applications are the normal practice. This scenario presents difficulties in selecting relevant exposure concentrations and application intervals and in interpreting results. The single application experimental design is easier to interpret in that it provides a clearly defined concentration-response relationship. In any case, simple addition of all multiple applications into a single dosing is probably inappropriate.

10. Does the SAP have recommendations regarding population and community models which might be most suitable for regulatory evaluations?

The Panel felt that population and community models should be restricted to Levels 3 and 4 because they require additional expertise and data that are not normally available at the Level 1 analysis. Models that focus on changes in species dominance, richness, and similarity are of value. These metrics can be compared among treatment levels in a micro- or mesocosm test and used in a 3rd or 4th level risk assessment. The appropriate model structure may be a canonical correlation of the set of biological, chemical and physical values. Because the metrics used (species richness, dominance, similarity) have high sampling variances, the associated variance in model output will also likely be high. It was suggested that models should focus on overall change in species dominance, richness, and similarity among treatment level are of value.

One Panel member identified several models that may be of use. They include RAMAS models which have been used for fisheries management. RAMAS software are available and easy to use. In addition RAMAS models are also probabilistic. At the community level, AQUATOX, CASM, and SWACOM are available, although they are not easy to use. CASM and SWACOM are probabilistic but have a narrower range of applicability (i.e., dimictic lakes) than does AQUATOX (rivers, reservoirs and lakes). Species sensitivity distributions can also be used as a simplistic tool for estimating community level effects. This tool, however, ignores the consequences of indirect effects.

11. Does the SAP agree with considering regional evaluations at Level 3, and focusing on 90% crop/use scenarios at Level 2, in order to direct initial evaluations toward high-end risk sites? Please discuss.

Panel members agreed that standardized high exposure scenarios (90 % crop use) were appropriate at Level 2 with regional assessments based on more realistic assumptions reserved for Level 3. This was qualified by one Panel member who felt that if the intent of ECOFRAM was to hypothesize test and to guide Level 3 tests, then regional assessments should be included in Level 2.

The Agency's document is not clear on how a 90% crop/use scenario is identified. Rather than rely on a percentile that can be interpreted in many different ways, the Agency should rely on identifying a range of high exposure scenarios based on use patterns, fate, etc., in problem formulation and then estimating risks for each scenario in Level 2.

The reasonableness of the 90% criterion depends on the reliability of the data bases used in the modelling and the broad applicability of the various scenarios. There may be a need to evaluate urban/agriculture use scenarios when the use patterns warrant it.

Before the model structure can be finalized, case studies are needed to validate the tiered risk analysis process. The modelling reserved for Level 2 (i.e., 90 % crop use) may be too conservative as to exposures, end points and species, making a third tier examination necessary in most cases.

Uncertainties in exposure estimates may remain high in regional assessments proposed for Level 3. The hydrologic component of the PRZM/EXAMS model treats multiple field plots over whole watersheds as independent, uncoupled simple 1-dimensional flow systems. However, these field plots are coupled hydrologic systems that exhibit complex 3-dimensional water flow and pesticide transport. Thus, it is difficult to assess the value of the information derived at Level 3 when hydrological processes that drive pesticide fate and transport are not being considered.

12. In reference to species sensitivity distributions,

A. How many species within a taxa should be tested at Level 3 to adequately characterize a species sensitivity distribution without the need for extrapolation factors?

It was agreed that the number of species necessary to remove extrapolation factors depends on the percentage of species that are to be protected and the sensitivity of the test species. The Panel concluded that by testing a fish, a *Daphnid* and an alga, one would expect to be within an order of magnitude or less of estimating the most sensitive species on an acute basis. If this same thinking were applied within a taxon, one would expect to characterize the sensitivity quite well with three species, assuming sensitive species were used (variability within taxa is less than across taxa). For example, aquatic insects, mayflies, stoneflies and caddisflies are frequently

used to assess chemical toxicity. Within aquatic zooplankton community, multiple *Daphnid* and Aamphipod species are often used. Evaluation of within taxa has received less attention than across taxa because it is recognized that sensitivity differences within taxa are less than across taxa, and there is greater interest in defining overall sensitivity in order to protect the entire ecosystem.

One Panel member noted that in the most simplistic description of the problem, the minimum number of species equals 1/(fraction unprotected). To protect 95% of organisms, the minimum number of species to test would be 1/0.05 = 20. It also was noted that when modes of action are known, testing can focus on the most sensitive taxa, in which case 6 to 8 species can be used to develop a reliable risk assessment. A recommended alternate approach for compounds with unknown modes of action included testing three species per family (18 total species) and applying extrapolation factors to distributions within a taxa to identify sensitive taxa. This would be done to define the taxa to test further. By testing five additional organisms from the most sensitive taxa (for a total of 8 species within that taxa), the criteria for effects distributions will be met. It should be noted that the latter approach requires testing of 21-23 total species which also meets the minimum criteria for evaluations of all taxa. Therefore, neither extrapolation nor safety factors would be required.

B. How should the aquatic taxa be grouped for evaluating species sensitivity distributions?

Generic recommendations for separation between taxa are difficult. The ability to separate and group taxa for evaluating species sensitivity is limited by the number of organisms within various phylogenetic groups for which there are standard toxicity tests. The classical separations that have been used and which still seem to apply are the following:

Fish – freshwater and marine (cold water and warm water species)

Invertebrates - (freshwater, e.g., zooplankton, insects) / (marine)

Plants - (algae and macrophytes)

There may be justification for the continued use of these broad categories. Herbicides are often toxic to both macrophytes and algae. Insecticides developed for terrestrial application often are toxic to a wide range of aquatic insects and zooplankton such as daphnids, crustacea, copepods and rotifers. Some of the fungicides and pyrethroid insecticides are highly toxic to a wide variety of fishes. However, as an alternative approach, a Panel members suggested groupings should be emphasized by foraging behavior, ecosystem(s) inhabited, and physiology.

13. In reference to additional sublethal effects testing,

A. Does the SAP concur that additional sublethal effects testing (such as immunocompetence) at Level 3, when a specific mode of action is of concern, will

improve risk evaluations? Please provide rationale.

There was agreement that additional sub-lethal effects testing may be useful. It can be postulated that studies based on known mechanisms of toxicity of chemicals at the lowest level possible in the screening would enhance the predictive capability of the risk assessment process. Even if the major mechanism of action of the chemicals is unclear, a non-invasive suite of screens may still be useful in identifying behavioral, reproductive and genetic levels of action.

One Panel member qualified support for such testing. Endpoints that are hard to extrapolate to the population or higher effects levels (e.g. enzyme testing) should be avoided. Survival and growth of different life stages and reproductive fecundity were identified as the most useful endpoints for estimating risks at the individual and population levels of organization.

B. Are there specific sublethal effects that the SAP would recommend as most useful to assess? Of particular interest are sublethal effects that may affect endpoints potentially suitable for regulatory decision-making (e.g., survival, fecundity).

A relatively new methodology is micro-DNA array screening (Winzler et al, 1999). It will be soon be available for chicken, mouse, and rainbow trout sequenced genes. In this procedure, mRNAs from animal cells or tissues are obtained after treatment with a chemical or chemicals under study and used to rapidly identify responsive genes that have been activated. This approach and other emerging gene analysis technologies (Roth et al, 1999) are potentially powerful tools for risk assessment. More prosaic sub-lethal tests that can and are being applied in ecotoxicology laboratories include: ELISA assays for reproductive state using male and female hormones, cholinesterase assays of blood and tissues, vitellogenin assays, P450 and related EROD assays and tests for OPIDN of appropriate organophosphorus pesticides such as methamidophos and isophenphos (Francis et al 1985).

One Panel member recommended that a workgroup be formed to consider ways environmental risk assessment can take advantage of rapid advances in the understanding of genetic expression at the molecular level. Pesticides have been in use long enough to theoretically have caused inherited effects at the population level of aquatic and terrestrial organisms in addition to the well known phenomenon of insect resistance.

Strong emphasis was placed on the need to continue standardized testing protocols. A recent evaluation of cholinesterase data submitted to the Agency for setting RfD's revealed lack of consistency and major problems in the conditions of clinical assays including the lack of controls and blanks. Errors greater than 70 percent were observed (Wilson, 1999). Suggested criteria for inclusion of assays were: (1) end points specific to the action of the chemical, (2) accepted validations of the applicability and standards for the assay, (3) positive and negative controls (4) ability of many laboratories to perform the assays, and (5) quantitative output.

ADDITIONAL COMMENTS

Several Panel members expressed the need for continued model validation and field testing. A generalized approach would be to use registrant data for chemicals already under registration and/or which are undergoing reregistration. How effectively the 4-level risk assessment process performs should determine possible changes to the number and approach in the assessment levels.

The Panel expressed another concern with the term "level of refinement." The term “level of refinement” is used for each stage of assessment. The term "level of refinement" suggests a refinement of something more crude. However, when the new methodology is adopted, the first level of refinement is simply the first stage of assessment, or the zeroth level of refinement. The terms "Tier of Assessment", "Stage of Assessment", or "Level of Assessment" are perhaps more appropriate and avoid confusion.

There also were many comments that appear to be an imbalance between exposure and toxicity assessment. Exposure assessments are based on computed values using fate and transport models at two levels of complexity. Model outputs are compared to actual toxicity testing data. In this context, it was noted that the Agency is using more exposure modeling. The Panel noted that PRZM/EXAMS may not perform satisfactorily in estimating concentrations of chemicals in certain watershed scale systems. Discussions at previous FIFRA SAP meetings suggested that the models are not sufficiently sophisticated to account for factors that may greatly influence exposure potential (e.g., presence of drainage lines, percent crop area, etc.), and that the current models do not account for groundwater discharge to streams or ponds, which are particularly important during dry periods. The Panel was pleased with the Agency’s use of monitoring data and its application toward the use of regression-based models to estimate pesticide concentrations. The Panel also supports the partnership with the U.S. Geological Survey which collects and maintains extensive water quality data sets (NAWQA) and has developed regression-based models such as SPARROW.

Finally, many Panel members encouraged the Agency to increase "data-mining" efforts. In the case of chronic and acute toxicity data, such efforts would help answer questions about the need for and the order of magnitude of extrapolation factors at all levels of biological organization. Particular concern was expressed about models that include more than one extrapolation factor. When factors are multiplied, the degree of conservatism in outputs has the potential to provide a strongly biased estimate of toxic potential.

PART 2:TERRESTRIAL TECHNICAL PROGRESS REPORT

DETAILED RESPONSE TO THE CHARGE

Questions on Terrestrial Technical Progress Report

1. The Terrestrial ECOFRAM Workgroup recommended that the LC_{50} test be modified by calculating an incipient LC_{50}, defined by the point on the study when the LC_{50} does not decrease by more than 1% over two days. (This modification was proposed through OECD as described in background Document #4.) Can the SAP comment on the utility of the recommended LC_{50} test modification to account for exposure durations for pesticides with moderate to long residue dissipation rates in wildlife foods, soil, and drinking water sources?

Incipient LC_{50} determination, as described by ECOFRAM, provides a good mechanism for evaluating the duration of adverse effects. The proposed design is particularly important in the case of compounds with long half lives or for transformation products with long half lives. Such data will demonstrate the magnitude and duration of toxicity that produces mortality. These LC_{50} data and environmental half lives can be used to estimate exposure potential for organisms inhabiting areas treated with the test chemical.

The actual percentage change in toxicosis over a 48 hour period should be assessed. For example, to determine a 10% change in toxicity, at least 10 organisms must survive at the point of diminished toxicosis. At doses higher than using an LC_{75} as the example, at least 40 animals would be required in the dose group to allow 10 to survive for determination of incipient toxicity.

There also is the danger of testing too few doses, since the regulatory processes attempt to limit exposure risks on the lower tails of the distribution. For example, the incipient LC_5 may be only slightly different from the standard LC_5, while the incipient LC_{50} may be significantly larger than the standard LC_{50}. Of course this example is only illustrative as the responses could be reversed.

One Panel member supported the idea of estimating incipient LC_{50} for pesticides with moderate to long persistence. Current LC_{50} estimates give no indication as to whether longer exposure duration would result in lower statistical estimates. Time to incipient LC_{50} should be reported. Defining incipient LC_{50} as the point in the study where decreases in LC_{50} are less than or equal to 10% over two days is an acceptable approach, but may be difficult to apply in routine tests. It is assumed that a statistical model is going to be used to decide when the endpoint has been reached. Otherwise, numerous organisms in each treatment replicate would be required (>40) with daily observations in order to detect changes in LC_{50} of 10% or less with confidence. It is also possible that no deaths may occur for periods of several days or more, subsequently followed by more deaths. A statistical model would take advantage of the entire time series data set. The 10% criterion would possibly be sensitive to choice of model. An alternative approach would be to use a segmented regression model. The model could be set to have a segment with a negative slope initially (as LC_{50} falls with time) and a slope=0 segment (i.e., no further decline in

LC_{50} with time). The segmented model is constrained to have the two segments meet at a "join" point. The "join" point represents time to incipient LC_{50}. The lack of independence between data points over time (common to many time series analyses) would need to be considered in developing an appropriate statistical modeling approach.

2. EFED has proposed two approaches for discerning appropriate exposure duration windows for calculating cumulative or time-weighted average exposures for short- and medium-term lethality risk assessments. These are (1) setting the window to match toxicity test duration and (2) extending the window over a protracted period limited to some minimal effect point.

> **A. Can the SAP express a preference for either approach and provide a rationale?**
> **B. If not, can the SAP define a more appropriate approach that utilizes existing data sets (i.e., without resorting to additional data requirements for pharmacological/pharmacokinetics data)?**
> **C. Should exposures (in the absence of additional data) be calculated as averages or cumulative over the assigned exposure window? Please provide the rationale.**

When conducting a Level 1 assessment, it is recommended that the exposure window be set to match the duration of the toxicity and that reasonable maximum (near peak) exposure be used to derive the EEC. In higher Levels of Refinement (Levels 2-4), it is recommended that the exposure window also be set to match the duration of the toxicity test (or, for example, the duration of the appropriate life stage of the organism) and that the exposure window be set to match the co-occurrence of the organism and the exposure i.e., if the pesticide is used in February, the organism and/or the food source that might be affected needs to be present in February. Further, in Levels 2, 3, and 4, it is recommended that a time weighted average be used. The rationale for this is that without further data, one would assume the exposure is fairly constant across short and medium term lethality risk assessments and hence, an average value would best represent the exposure profile. Additionally, because mortality is the endpoint being assessed, one would not expect this endpoint to be as sensitive to exposure fluctuations as development in a critical life stage.

One Panel member expressed uncertainty concerning what the Agency is implying by the second option. When calculations lead to significant extrapolations beyond the duration of the acute toxicity test, there will be high uncertainty. In general, the Panel member supported the option of matching exposure and toxicity test duration in Level 2 or higher assessments. However, an exception may be necessary in cases where toxicity test duration was not sufficient to reach a further effect. Exposure information (e.g., field dissipation rates, exposure medium) also needs to be considered. In Level 1, the Agency may want to consider using peak exposures to be appropriately conservative.

Perhaps an even better option would be to set exposure duration to match time to incipient effect. Where existing data do not permit matching of exposure and test duration (e.g., persistent pesticide, but chronic tests unavailable), one could estimate cumulative effect doses

causing effects in toxicity tests and compare such doses to cumulative exposure doses based on expected chronic exposures in the field. Cumulative exposure doses should be adjusted to account for changes in residue concentrations over time as well as organism depuration and metabolism, and toxicity of metabolites.

Average residues are reasonable for persistent pesticides, but not for prey residue concentrations which decline slowly following application. Cumulative approaches may be necessary where average dose would be highly variable over time. Estimating cumulative doses can be difficult because of declining residues in the field over time, and organism movement from day to day. Random walk models could be used to address organism movement over time.

Exposure duration is largely dependent on field dissipation rates. The American Chemical Society is holding their Fall meeting on August 20-24, 2000, in Washington, DC. A symposium has been organized to address field dissipation. Participation of the Agency staff in this Field Dissipation forum will allow access to data from new and benchmark dissipation studies.

3. Because of the present lack of avian inhalation and dermal toxicity data, EFED has proposed an equivalency factor approach, based on laboratory rodent acute toxicity potency comparisons across exposure routes, to normalize exposures from dermal and inhalation routes for birds.

A. Does the SAP believe that, in the absence of specific pharmacological/pharmacokinetics data, that this approach is reasonable and if so, why?

B. If not, can the SAP provide insight into quantitative methods, using the existing data sets, to facilitate a comparison of different exposure routes to the existing toxicity data presently required for birds?

C. There is a recognized need, in higher Level of Refinement assessments, for additional toxicity data for routes of exposure other than oral to reduce extrapolation uncertainties. Can the SAP provide any detailed guidance on developing testing protocols (technical points to consider or knowledge of existing methods) for avian effects in birds via the dermal or inhalation routes?

Inhalation rates for a given bird are not static and may represent another probabilistic term. The suggested 3X factor to increase inhalation based on high metabolic rates should also be modeled probabilistically. Evaluation of the avian activity during tests already in the literature could also help evaluate the need for this 3X factor. After determining/evaluating the uptake rates for inhalation, sensitivity analysis should be done to determine the need to model this parameter. There have been very few studies of dermal or inhalation toxicity of birds. If inhalation and dermal exposures are significant routes and vary significantly based on metabolic activity, more data to describe these routes of exposure are essential and research in these areas should be

prioritized.

There are many problems with automatically accepting statements and approaches used by the scientific community in studying dermal absorption of mammals as applicable to birds. One problem is that many experiments of mammals study the permeability of and modeling dead, not living skin. The results described may be dependent on the passive properties of diffusion of the tissue rather than possible metabolic transformations and active movement of the molecules. The events that occur across the living skin are often understudied. Second, avian skin is different histologically from mammalian skin and findings on physiologically living skin may not be transferable from mammal to bird. There have been few studies of permeability of the skin of birds. Investigators may circumvent the issue by applying agents to the comb (if the subject is a chicken) or under the wings to avoid plucking feathers. Much work is needed in this area. In one case, results indicated that cholinesterase activity in the blood of pigeons and kestrels remained depressed for weeks longer when an organophosphate agent was applied to the feet (and the feet washed after application) than when it was given orally (Bartkowiak and Wilson, 1995).

In the absence of pharmacological/pharmacokinetic data, should an equivalency factor approach using rodent toxicity data be used to account for potency differences across exposure routes of birds? The approach seems reasonable if pesticide potencies are expected to differ substantially across exposure routes. However, the Panel was not certain how critical this issue is. Presumably, this approach could not be used to address the preening exposure route. Perhaps the only other possible approach would be to develop generic equivalency factors. This would involve calculating exposure route equivalency factors using results of pesticide toxicity tests, where tests of different exposure routes have been done for birds. Conservative equivalency factors (e.g., 95th percentiles) could be used for Level 1 assessments; entire distributions could be used in higher level assessments.

4. Published studies suggest that for some pesticides and birds, exposure via preening can be an important route. Can the SAP comment on the need for considering this route of exposure? If the route is believed to be appropriate for consideration, is the SAP familiar with any quantitative methods for including this route into the overall exposure assessment?

At this point in time, the Panel believes it is premature to include preening as a routine test of avian exposure to pesticides. The importance of preening, percutaneous, and inhalation routes are vital parts of total exposure; especially from pesticide spray application. The Panel recommends that each of these routes be carefully studied for a better understanding of their proportional contributions to exposure. Initially, this will require development of sound protocols for evaluation of preening subsequent to several application techniques including at least representative wettable powders, emulsions, flowables, dusts, and technical grade (ULV). These tests would allow focus on the more potentially hazardous end-use products. Perhaps such baseline testing with one species each of duck (mallard), quail (northern bobwhite), and small passerine would provide evidence of a common pattern of proportional contributions that may have broader use in hazard prediction. More likely, this initial testing would increase uncertainty

and necessitate additional work with other species and behaviors. If the Agency chooses to pursue preening as a special route of exposure, it must also consider the potential hazard of grooming in carefully selected mammals for comparability to results for laboratory animals.

Clearly, when wildlife are directly sprayed, all routes of exposure contribute to their response, but most often ingestion of contaminated food and water is the main route of concern in incidents of acute exposure. In these cases, birds and mammals are often severely over-dosed and any additional effects of ingestion from preening may be inconsequential. However, in the event of wildlife moving about in contaminated habitat, preening could become uniquely important as a temporal event of repeated exposures. This type of situation would require special testing if the purpose is to separate ingestion from preening and foraging compared to dermal exposure. This distinction is probably not of any practical importance. Even in this situation, foraging and hydration would likely remain more important as the principal source of exposure. Perhaps even more important to birds than unique exposure from preening is the potential for brooding adults to contaminate eggs. This route of exposure has proven embryotoxic in the laboratory, especially for pesticides in an oil solvent.

The importance and resolution of multiple routes of exposure including oral (foraging, hydration and preening), dermal, and inhalation, must be addressed to legitimize pesticide risk assessment for wildlife. The Panel concurs that this topic is poorly understood by the scientific community. The Agency might consider conducting a workshop on routes of exposure with an emphasis on pesticide application techniques and formulated end-use products.

5. EFED is trying to determine the best basis for acute effects characterization in the risk assessment process. Can the SAP provide guidance on selection of either the acute single oral dose or the dietary toxicity study as the basis for acute lethality assessments?

The Panel complimented the Agency on its terrestrial toxicity summaries. The topic of LD_{50s} and LC_{50s} is controversial. The LD_{50} is a basic toxicology test that yields relative potencies of chemicals and, with due considerations, sensitivities between species. An excellent discussion of LD_{50} comparisons is found in a recent review of Mineau et al. (2000, In Press). LC_{50s} are 5 day exposure studies of young animals involving elements of feeding trials, growth and accumulative toxicity. The test has been criticized by some scientists including Mineau and Baril (unfortunately, the paper quoted by the Agency was an unpublished report and not available to the Panel.) Another Mineau report (1996) persuasively argues that the dietary studies are convoluted including acute toxicity, nutritional and behavioral affects that are reflected in the growth of the young birds. A well done acute toxicity study coupled with carefully designed subchronic tests might provide more reliable information than dietary studies that may confuse direct toxicity with behavior and nutritional problems. One issue is the ability of the test(s) to span a sufficient dose range to encompass both tails of the dose/response curve. However, one Panel member stated that the LC_{50} can be most informative when carefully performed.

There are many technical considerations in the performance of oral tests. One is delivery

of the chemical, whether by gavage or capsule. To be certain of the oral dose delivered to birds, one recommendation is that the treatment be undertaken after withdrawing food overnight to empty the crop. Lack of awareness of this in the past has led to widely differing reports of toxicity of the same chemical such as isofenphos (Chow et al, 1986).

Another factor is the rapid growth and maturation of young birds. Birds grow so rapidly that disturbances in vitamin availability are reflected in a matter of hours (e.g., Gries et al, 1972). Additionally, and little recognized, are the major changes in the neuromuscular system that occur after hatching. These include the establishment of the adult pattern of enervation of fast twitch and slow tonic muscles and the major effect that anticholinergic chemicals can have on skeletal muscle itself by damaging regions adjacent to motor end plates due to excess acetylcholine (Wecker and Dettbarn, 1976; Leonard and Salpeter, 1979).

Another issue is field validation. The benchmark screenings and the mathematical statistics that have been developed to accompany them unfortunately raise the question of the reliability of the data input themselves and their ability to predict real world outcomes. The better the probabilistic models, the more important it is to subject them to careful field validation.

A positive feature of LC_{50} dietary studies is that they may yield information on food consumption, weight loss, and aversion behaviors that cannot be gained from an acute oral toxicity experiment. However, it is expected that problems in variability will arise if the intention is to use existing benchmark data for the early tiers in the process. One suggestion is to consider when new species are introduced into the tiers. The risk assessor's job may be made simpler if the screening is done starting from LD_{50}s, and food ingestion enters later in the refinements.

One Panel member noted that dietary toxicity studies would be preferable because the exposure route better approximates dietary exposures in the field, particularly for pesticides with low to moderate field dissipation rates. The single dose test may be more appropriate when estimating risks following gorging, or in situations where exposure is expected to fall rapidly following application. One Panel member proposed a solution to this problem. One recommended procedure is scoping trials, working with pairs of birds to choose dose ranges for physiological and biochemical studies, a practice that is both efficient and consistent with animal welfare concerns. Perhaps such preliminary tests can be incorporated in a regime that keeps both LD_{50} and LC_{50} tests.

The Panel was reluctant to recommend more research even though it seems clear no single test or species seems to fit all situations or chemicals. Nevertheless, the Panel discussions moved towards requiring new research on new species and conditions. Hopefully the recommendations provided by the Panel will help reduce the amount of new research that may be necessary to fit the new probabilistic models and will help sort out reliable existing data bases. Regardless, there is no substitute for careful standardization of testing and acceptance of a test only after validations of its input and output.

6. In the opinion of the SAP, what are the minimum study requirements for each Level of Refinement for both lethal and reproductive effects? How many species should be tested at each Level of Refinement to adequately characterize a species sensitivity distribution?

Although there are often differences in toxicity to species within each trophic level, it is generally possible to describe a range of sensitivities according to the mechanism of toxicity and relationship of metabolic processes to feeding habits and feeding guilds. The time, resources and animals, would outweigh the benefit of such information. However, one practical method that could be used to determine the lethal and sub-lethal effects should include testing of birds (and mammals) within each representative trophic levels. The first level of refinement can be based on characteristics associated specifically with trophic levels, ideally using a representative species from each of the primary trophic levels. This would require considerable additional research since there are few data on the toxicity of chemicals to raptors and higher trophic level birds and mammals. This will likely cover a large number of exposure situations and cover an adequate range of metabolic and physiologic based sensitivities to a chemical. Although this should cover most exposure scenarios, there will always be outliers that do not fall within the expected toxicity responses of the tested species. This challenge should be accepted and it should be assumed that the most appropriate surrogates in each trophic level will provide the best first level of risk estimate. The use of risk assessments in this approach will be a valuable tool to make the necessary extrapolations between species when necessary.

The endpoints used for reproductive effects are generally more complex than those used to evaluate lethal effects. The Panel recommends that the most important endpoint in the data set has been the impact of egg sequestration of persistent chemicals and the ability of some species to depurate the chemical without significant primary effects. In the progression to more refined risk assessments, the most critical parameters include the bioavailability of the chemical and the overall distribution in the environment. To refine the reproductive testing scenario, however, the most productive addition to the avian test protocols would be to track and evaluate feed consumption of the hens and to provide a paired feeding group that follows the treated birds by a few days or week to evaluate the important effects of food consumption on egg production.

Based on ECOFRAM, at Level 1, two species are recommended. With these data, all extrapolation factors are below 5 and the Coefficient of Variation about the 5th percentile LD50 extrapolation is less than 8%. Therefore, by testing a second species, the Agency should be able to eliminate the uncertainty factors at this level. Based on the extrapolation factors presented in ECOFRAM and possibly recent peer reviewed publications of similar data information, there seems to be little to be gained for moving past three species. Also, diminishing returns at three test species is similar to the information presented by one Panel member that indicates three aquatic species represent most sensitive species for 88% of tested compounds. Also, moving rapidly to non-avian screens are essential.

It was stated that it is unrealistic to expect that species sensitivity distributions can be developed for birds for the vast majority of pesticides that will be assessed by the Agency. In

general, one prefers an approach that uses information on pesticide use patterns, properties, and fate to select focal bird species that are most likely to receive high exposures (e.g., kingfishers for pesticides that bioaccumulate through aquatic food webs). The goal of toxicity testing at higher levels of refinement would be to collect better information on pesticide effects to the focal species (e.g., testing of focal species or close surrogate, testing involving different exposure routes and duration, testing for different endpoints such as growth, fecundity, avoidance, etc).

Testing many bird species at higher levels of refinement would also divert resources from considering risks to mammals, reptiles and other terrestrial species — the latter need to be given higher priority for testing.

7. EFED has recommended modifications to the avian reproduction test to provide dose-response information for sensitive endpoints. Can the SAP provide suggestions for protocol design and/or the most important considerations that should be factored into the design of the study?

The review prepared by Bennett and Ganio (Overview of methods for Evaluating Effects of Pesticides On Reproduction In Birds) nicely summarizes the issues associated with conducting reproductive studies. The Agency is currently participating in on-going efforts by world experts under OECD to develop a standard harmonized avian reproductive study. It seems prudent to adopt the recommendations from this expert body.

Protocols should be modified to facilitate use of regression analysis to develop dose-response relationships (more doses, more organisms per replicate at low doses, less replicates). Guidance would be needed on spacing of treatments, choice of model (suggest Generalized Linear Modeling framework), assessing goodness-of-fit, etc.

The future direction of reproductive testing diverts from the ANOVA approach and associated designs. The Panel supports the use of regression analysis for assessing dose-response and making estimates of risk when exposure data are available. Additionally, future tests should place less emphasis on mortality and growth and greater emphasis on birth rate per female and reproductive recruitment into the population. Assessing pre-recruitment deaths (embryo mortality and post hatching exposure) should become the focus of the study.

With ANOVA test design, the determination of whether the independent treatment variable affects the dependent response variable will depend on the ratio of variability between treatments versus within treatments. If there is high variability within treatments and only one or two replicates per treatment, changes between treatments may not be detected as statistically significant, regardless of treatment number. Whether one should add more treatments or more replicates will depend on how variable the response is; i.e., how precisely one needs to be able to estimate each treatment mean so that differences between treatment means will be apparent. To that end, efforts are being made to reduce the variance in data that has been observed in reproduction tests to date.

One must consider the variance of the estimate of most interest to determine how to optimize a design. Many of the variance estimates in regression have variance of the X variable in the denominator so that maximizing the variance of X will minimize the variance of the estimator (intercept, slope, predicted Y, etc). Additional X levels provides more information about the true shape of the response curve. Equal replication at each level of X gives the greatest power. Placing one's mean X value where there is the most interest in the response gives minimum variance for that level of the response.

When regression data are obtained by experiment, the levels of X at which observations on Y are to be taken are under the control of the experimenter. Among other things, the experimenter will have to consider the following, as stated by Neter et al. (1990):

1. How many levels of X should be investigated?
2. What shall the two extreme levels be?
3. How shall the other levels of X, if any, be spaced?
4. How many observations should be taken at each level of X?

There is no single answer to these questions, since different purposes of the regression analysis lead to different answers. The main objective may be to estimate the slope of the regression line, or in some cases to estimate the intercept. In many cases, the main objective is to predict one or more new observations or to estimate one or more mean responses. When the regression function is curvilinear, the main objective may be to locate the maximum or minimum mean response.

To illustrate how the purpose affects the design, consider the variances of the slope, the intercept, a predicted Y value, a prediction of the value of Y for a new observation. The variance of the slope is minimized if the variance of X is maximized. This is accomplished by using two levels of X, at the two extremes for the scope of the model, and placing half of the observations at each of the two levels. Of course, if one were not sure of the linearity of the regression function, one would be hesitant to use only two levels since they would provide no information about possible departures from linearity.

If the main purpose is to estimate the intercept, the number and placement of the levels does not matter as long as the mean of the X values equals zero. On the other hand, to estimate the mean response or predict a new observation at a particular X value, it is best to use levels of X so that mean of X is the X value of interest. If a number of mean responses are to be estimated or a number of new observations are to be predicted, it would be best to spread out the X levels so that mean of X is in the center of the X levels of interest.

Although the number and spacing of X levels depends very much on the major purpose of the regression analysis, some general advice should be given, at least to be used as a point of departure.

Two levels should be used when the object is primarily to examine whether the

independent variable has an effect and in which direction that effect is. Use three levels whenever a description of the response curve by its slope and curvature is likely to be adequate; this should cover most cases. Use four levels if further examination of the shape of the response curve is important. Use more than four levels (5-8, for example) when it is required to estimate the detailed shape of the response curve, or when the curve is expected to be asymptotic, or in general to show features not adequately described by slope and curvature. Except in these last cases, it is generally satisfactory to use equally spaced levels (in log space) with equal numbers of observations per level (Cox, 1958). In the latter case, which is often encountered in chronic reproduction studies, the doses should be closer together at the low dose part of the curve (in arithmetic space). Properly setting the doses to obtain a well defined dose-response curve usually requires a preliminary study. Additionally, study designs have been used where the number of test animals is increased in the treatment levels of greatest interest (lower treatment levels) in order to increase the ability to perform low-dose extrapolation.

Finally, the Panel commented that it would be useful to consider a variety of fecundity endpoints (e.g., percent successful mating, embryo toxicity, hatching, survival and growth to different times following hatch) in the protocol design.

8. EFED has proposed options for interspecies extrapolation factors for both lethal and reproduction risk assessments.

A. Can the SAP comment on whether the proposed method is appropriate?

B. Can the SAP recommend an alternative or additional approach, with supporting detail?

C. Is the proposed approach for reproduction effects appropriately conservative under the existing limitations of available data, and if so, why?

D. If not, does the SAP have a preference for an approach for reproduction effects extrapolations at this time?

There are difficulties in extrapolating from one species to another, not the least of which is the fact that the data bases are mostly acute LD_{50} studies. The discussion of extrapolation factors in the Terrestrial System Report was reminiscent of discussions of deterministic uncertainty factors. The Panel suggests use of a review that is in press and has been seen by the Agency (Mineau et al, 2000). The approach seems reasonable as described. However, several recent papers raise the important problem of scaling factors (Mineau et al, 1996; Sample and Aranal, 1999). The weight of evidence is that the traditional $2/3^{rd}$ power corrections derived from mammal studies and surface/volume considerations are not accurate and may differ for small and large birds. Such considerations indicate that more work is needed in this important area. One alternative for the risk assessor, is to rely on extrapolation (in this case uncertainty) factors of multiples of ten. Unfortunately, this would be a giant step backward. It would be better to support development of extrapolation factors for size classes of birds and other distinctions as necessary and achieve a consensus among avian scientists and risk assessors as to which ones to apply. In addition, given the near total absence of supporting data, the factor suggested (10) is somewhat arbitrary. A larger factor (100?) may be justified for Level 1 until data become available that justify lowering of the factor. No factors should be used in higher level

assessments. The database for this extrapolation factor could be expanded by considering data of industrial chemicals. Otherwise, little can be done until better data are developed.

The Panel agreed that this an appropriate analysis as developed by ECOFRAM. During the implementation of this extrapolation factor, the use of a passerines (e.g. red-wing black bird) would seem prudent because passerines represent sensitive species in many circumstances. By using the empirically derived extrapolation factors, the need for uncertainty factors is greatly diminished to the point of no uncertainty factor at 4 tested species.

Body size plays a very small role in the overall determination of the dose of a pesticide and/or toxicity. One Panel member noted that correcting for body size may not be necessary. Many pesticides in the Baril and Mineau (1996) study showed no significant relationship between sensitivity and body size. A point estimate extrapolation factor should not be used at Level 2 or higher. Any given untested bird species may be less, equally or more sensitive than tested species. Thus, in the absence of toxicity data for an untested species, a distribution should be used to represent variability in species sensitivity. Using a conservative point estimate in Level 2 defeats the point of conducting a probabilistic risk assessment. However, such a point estimate is appropriate in Level 1. No alternative approach can be recommended at this time. Metabolic rate is a fundamental parameter in the relationship of toxicity to species and metabolic rate is a function of physiological activity (oxygen consumption and respiratory rate) per unit active cells. There are few data that directly relate metabolic rate to toxicity in birds and mammals. This is an area of research that could prove useful and directly applicable to pesticide toxicity.

9. Can the SAP provide guidance on what additional species would be the most appropriate for testing for both lethal and reproductive effects at the higher Levels of Refinement?

Regardless of the level of refinement, risk assessment of birds needs to be concerned with different life histories and life styles. Major divisions among birds include altricial and precocial, determinate and indeterminate layers, resident and migrant species. In addition there are diversities in food gathering such as raptorial, seed eating, fruit eating, and fish eating. Proposed uses of the chemical under study and the life style (ecological niche) of the birds are important matters in the selection of which species should be examined. Another complementary approach that should be considered is to use the chemicals themselves to help in the choice of species. The data base of chemicals structures used in pesticides and the effects they generate must be large. Application of QSAR (Quantitative Structure Activity Relationships) to analyze the chemical structures could give insight into the expected modes of action, suggesting what parameters and species might be best suited for the higher tier studies. One proponent of this approach is Corwin (1997) who is constructing a database of QSAR for all types of reactions. In 1997 the database contained over 10,600 examples, of which 4,000 were from biological systems.

One concern is the lack of other species at lower levels of refinement. Risks that might pertain to other species, life styles and biology might not be revealed at the lower tiers with only bobwhite and mallards. Altricial birds differ enough in their life cycle to make extrapolation to precocial species risky. Raptors such as the American Kestrels should also be considered under

special circumstances. Kestrels are captive bred in a few facilities in the U.S. and Canada and are readily trapped or convinced to use nest boxes in the wild. The advantage of this species is the food chain effects and egg shell thinning that might be revealed. A disadvantage of this approach is the expense.

Ideally, species tested should correspond to species at greatest risk. Practically speaking, this would be difficult because standardized protocols are not available for the vast majority of bird species. One alternative is to develop protocols for representatives from broad feeding classes (herbivore, granivore, insectivore, omnivore, predator) and/or bird taxa groups. For specific pesticide assessments, testing could be limited to the feeding class or taxa group most at risk (i.e., the focal species). Once focal species are identified, tests considering different endpoints (lethality, reproduction) and routes of exposure (dietary, dermal, preening, etc) should be considered.

10. The Terrestrial Technical Progress Report presents methods for predicting dose-response slopes for extrapolated sensitive species.

A. Does the SAP have suggestions on ways to improve these methods or can an alternative approach be taken?

B. Can the SAP suggest methods for similar extrapolations for reproduction effects?

When only one dose-response has been measured, an estimate of the mean response is available, but an external estimate of the variability must be used. The approach of the Terrestrial Workgroup to use a coefficient of variation based on sampled data across species for other chemicals seems to be reasonable. The Panel suggests using the coefficient of variation across species rather than the standard error of the estimate (p.48, 1c of the *Terrestrial Technical Progress Report*) for the single dose-response as the measure of variability because the standard error of the estimate does not have a component of across-species variability.

When the dose-response has been measured for several species, using the mean of the newly measured slopes as the mean of the distribution of slopes seems reasonable. A measure of variability is still needed. Continuing to use the coefficient of variation based on sampled data across species for other chemicals seems to be appropriate for estimating the variability. Using a uniform distribution between the extremes of the newly measured data in essence indicates that the focal species cannot be more sensitive than any of the measured species, an assumption that does not seem reasonable when using a small number of measured species.

The Agency could consider taking a Bayesian approach to estimating the variability. The coefficient of variation based on sampled data across species for other chemicals could be used in developing the prior estimate for the variability. Newly measured values would then be used to update the variability estimate. The utility to this approach is that it would work for two or more newly measured species without an arbitrary cutoff or switch between methods, and would support estimating other percentiles of the distribution.

A Bayesian approach also could be used to develop a joint estimate of the mean and

variance of the dose-response data. This approach would then replace previously mentioned approaches for mean and variability estimation. The Panel recommends use of the same statistical methods as for mortality effects.

The approach of selecting a conservative percentile (flatter) dose-response slope from the distribution of dose-response slopes seems reasonable for level 1. At higher levels and in the absence of information on the focal species, entire distribution of slopes should be used. The appropriate distribution for slopes would be a normal distribution.

Given the limited data available on slopes for reproductive endpoints, any approach used will have high uncertainty. The Panel suggests considering the use of a "generic" slope distribution based on all pesticides or the pesticide class to which the pesticide of interest belongs.

11. Can the SAP provide guidance on the methods for developing natural history information on avian species in agro-ecosystems that would be appropriate for use in probabilistic assessments? This would include information on avian census, time budget and dietary proportions on and off treated fields, available food sources, and others.

The best approach for developing natural history information on avian species in agricultural ecosystems is astute field observations, radio-tracking of key species, and ground-truthing of geographic information system (GIS) mapping. It is essential that wildlife movements in and around pesticide treatment areas be understood. For example, are birds observed flying over a treated area in transit or actually using the field? Often agricultural monocultures are not particularly attractive to many wildlife species during different stages of crop development and pesticidal application. For different regions, and especially in arid climates, irrigated croplands may be attractive to wildlife for hydration. Exposure of wildlife through hydration has long been considered a critical source of poisoning yet has rarely been evaluated.

The main hazard concerns from contemporary, comparatively labile pesticides are for localized acute exposures and reproductive effects from temporal low-grade repeated exposures. Radio-transmitters can be used to monitor movements around treated habitats including the agricultural field and edges. Transmitters will be especially useful in determination of movements from distant nesting sites to fields for foraging and hydration. There are case reports of birds moving many miles to forage and obtain food for nestlings. This has resulted in parents being poisoned and not returning to the nest resulting in nestling starvation, or parental delivery of contaminated forage to nestlings. It is possible to place a short-term ligature on the neck of nestlings for retrieval of forage and residue analysis. The most critical and rewarding tact is the location of nests and repeated observation of fledging success. Time-lapse photography may be considered. This may be coupled with temporal evaluation of biochemical changes through non-destructive blood sampling.

These basic behavioral and exposure techniques can be overlaid on studies of replicated application of the pesticidal end-use product. The field applications must be evaluated for a variety of treatment levels in order to develop the dose-response association critical to predictive

(probabilistic) toxicology. Initially, at least three fields (replicates) per treatment should be evaluated for control and three to five geometrically arranged end-product application rates. The highest rate should exceed the optimal crop treatment rate. Hazard assessment toxicology, whether in the laboratory or field, is a function of dose-dependent exposure (direct and temporal). With this information, the risk assessor can predict effects through interpolation, but must be extremely cautious when attempting extrapolation to different wildlife species, crops, and agricultural regions.

At each exposure level, different effects may occur, e.g., altered behavior and reproduction at low levels, direct mortality at high levels. It will be incumbent upon the investigator to define possible effects prior to the evaluation, and carefully observe for them over time. It is important to attempt to quantify wildlife use of treated fields, but it is especially important to attempt to study the edge habitat to include potential zones of spray drift. Radio-tracking of birds captured in the field will provide critical insights into amount and distance of movement.

This is an area where industry and the Agency could quickly fill a data gap. Peer review and government reports could be mined for data. More importantly, there is significant data available from past field studies. If a Cooperative Research and Development Agreement was established to compile this information and fill the data gaps, then a consistent and broad based data set would be available for use in exposure assessments. This cooperation is urgently needed to advance the risk assessment process.

For most variables, the 1993 EPA Wildlife Exposure Factors Handbook provides estimates for both centrality (mean, median) and variability (standard deviation, confidence limits). This information together with knowledge about the variable can be used to select and parameterize distributions. The Handbook provides citations for data sources which may be accessed when more information is required for specifying distributions. Key natural history variables not summarized in the Handbook include many foraging variables such as proportion of time foraging in fields, buffer areas, and outside agroecosystems. The Panel believes this type of information is limited. Studies on foraging behavior are needed. To facilitate development of foraging information, the Agency should develop a short list of species that would serve as representatives of broad feeding classes or taxonomic groups. This would avoid having to conduct foraging behavior studies for 100s of species — an impossible task.

Several large databases such as BASINS have a wealth of information that could be used together with GIS tools to consider issues such as availability of suitable habitat for avian species, size of buffer areas, proximity to water bodies, availability of migration corridors, etc. Where information is lacking on key natural history variables, conduct “what if” analyses to determine impacts on risk estimates (e.g., try 100% field exposure versus 100% buffer exposure).

Closing Question

1. What approach would the SAP recommend to move toward validating the risk

assessment processes presented in the Technical Progress Reports? Please provide specific recommendations.

The Panel unanimously agreed that more case studies need to be conducted. Perhaps the Agency should consider conducting assessments on safer pesticides and high risk pesticides through the system, to determine if they exit from the system at the expected appropriate level.

Verification of modeled toxicity, must be field validated. There are numerous examples of situations where sophisticated modeling incorrectly predicted chemical effects on wildlife. These errors can occur as false positives or false negatives. It would be unwise to require Level 2 or 3 risk assessments without requiring confirmatory field validation of the model predictions of exposure and more importantly effect.

Several Panel members also commented that there appeared to be numerous conservative assumptions and extrapolation factors with the Level 1 assessment. Thus, there was the perception that a Level 1 assessment may not be able to screen pesticides or uses of low concern. Thus, refinements may need to be made to the Level 1 assessment.

Use of other lines of evidence (e.g., incidence information, field observations following pesticide applications, biological monitoring) seems to have not been considered in the present approach for assessing pesticides. The Agency is strongly encouraged to consider other lines of evidence and develop guidance on how such information should be used for pesticide risk assessments.

The importance of problem formulation is not emphasized enough in the current document. Problem formulation should be used to direct assessments and generation of exposure and effects information at all levels, but particularly so at Level 2 or higher. It appears that the framework, as currently written, has overlooked an important group of animals - mammals associated with agroecosystems. The Panel is concerned that mammals are not considered in the implementation plan for probabilistic risk assessments of pesticides. As part of human health data requirements, data are submitted on acute and chronic toxicity of pesticides to rodents. This information could be used as the basis for characterizing effects to wild mammals. In some respects, assessments of mammalian species are easier than would be those for avian species (e.g., mammals are less mobile and natural history information is readily available). The assessments should always include exposure and risk information for mammals.

These data and information being incorporated into any risk assessment should be evaluated for scientific acceptability. In some instances, data used to generate a LOAEL or NOAEL cannot be defended statistically. In some cases, these data are presented as relevant endpoints for toxicity estimates. In all cases the gap between the NOAEL and LOAEL should be evaluated and used in the assessment to indicate some measure of confidence that the suggested endpoint is acceptable.

There are different issues relating to these data needed for risk assessments of new

chemicals versus older chemicals. New chemicals and especially new classes of agricultural chemicals will need considerably more base data than existing agricultural chemicals. The Panel suggests that the more important information for existing agricultural chemicals consist of incident information and reports from actual use scenarios, particularly when field effects are seen.

It is clear that many of the risk assessment parameters are more important in the final assessment and should be reported as a sensitivity analysis and a series of "what if" exposure scenarios. There is wealth of data within the various agricultural chemical companies that address the issues that were discussed at the meeting. In the best of worlds, these data would be in a form that could be used by all researchers and regulators to generate risk assessments for chemicals and for classes of chemicals. The Agency is encouraged to develop a means of providing these data to researchers and risk assessors.

REFERENCES

Bartkowiak, D.J.; & Wilson, B.W. Avian plasma carboxylesterase activity as a potential biomarker of organophosphate pesticide exposure. Environ. Toxicol. Chem. (1995) 14(12), 2149-53.

Best, Louis B.; Fischer, David L. Granular insecticides and birds: factors to be considered in understanding exposure and reducing risk**.** Environ. Toxicol. Chem. (1992), 11(10), 1495-508.

Brix, K.V., F.J. Dwyer, W.J. Adams, C.E. Ingersoll and D.K. DeForest. 1999. Evaluation of Relative Sensitivity of Threatened and Endangered Species to Contaminants. SETAC Annual Symposium, poster paper, abstract published in symposium abstract book, Society of Toxicology and Chemistry, Pensacola, FL.

Chow, Edward; Seiber, James N.; Wilson, Barry W. Isofenphos and an in vitro activation assay for delayed neuropathic potential. Toxicol. Appl. Pharmacol. (1986), 83(1), 178-83.

Cox, D.R. 1958. Planning of Experiments. New York: John Wiley and Sons, p. 141-142. Neter, J., Wasserman, W., and Kutner, M. 1990. Applied Linear Statistical Models, Regression, Analysis of Variance, and Experimental Designs. Third Edition. Irwin, Homewood, IL., Pp. 176-177.

Cross, R. H. M.. Features of structure and size distribution of collagen fibrils in avian dermis. Electron Microsc. 1998, Proc. Int. Congr., 14th (1998), 4 75-76. Publisher: Institute of Physics Publishing, Bristol,

Cunningham, Mary Jane; Zweiger, Gary B.; Panzer, Scott R.; Seilhamer, Jeffrey J. (Incyte Pharmaceuticals, Inc., USA). Toxicity testing using a microarray of nucleic acid toxicological response markers PCT Int. Appl. (2000), 76pp.

Francis, B. Magnus; Metcalf, Robert L.; Hansen, Larry G. Toxicity of organophosphorus esters to laying hens after oral and dermal administration. J. Environ. Sci. Health, Part B (1985), B20(1), 73-95.

Gries, Christian L.; Scott, M. L. Pathology of pyridoxine deficiency in chicks**.** J. Nutr. (1972), 102(10), 1259-67.

Hansch, Corwin; Gao, Hua. Comparative QSAR: Radical Reactions of Benzene Derivatives in Chemistry and Biology Chem. Rev. (Washington, D. C.) (1997), 97(8), 2995-3059.

Henderson, J. D.; Yamamoto, J. T.; Fry, D. M.; Seiber, J. N.; Wilson, B. W. Oral and dermal toxicity of organophosphate pesticides in the domestic pigeon (Columba livia). Bulletin of Environmental Contamination and Toxicology, v. 52, (1994.): 633-640.

Hill E.F. Acute and subacute toxicology in evaluation of pesticide hazard to avian wildlife. In "Wildlife toxicology and population modeling: integrated studies of ecosystems". Eds RJ Kendall and TE Laches, CAC Press 207-226 (1994)

Hudson, R.H., Tucker, R.K., and Haegele, M.A. 1984. Handbook of Toxicity of Pesticides to Wildlife, 2ed. U.S. Fish and Wildlife Service Resource Publication 153, Washington, DC.

Kimerle, R.A., A. F. Werner and W.J. Adams. 1984. Aquatic hazard evaluation principles applied to the development of water quality criteria, Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTM STP 854, R. D. Cardwell, R. Purdy, and R. C. Bahner, Eds., American Society for Testing and Materials, Philadelphia, 1985, pp.538-547.

Leonard, John P.; Salpeter, Miriam M. Agonist-induced myopathy at the neuromuscular junction is mediated by calcium. J. Cell Biol. (1979), 82(3), 811-19.

Menon, Gopinathan K.; Brown, Barbara E.; Elias, Peter M.. Avian epidermal differentiation: role of lipids in permeability barrier formation. Tissue Cell (1986), 18(1), 71-82.

Mineau, P; Baril A; Collins BT; Duffe J; Joerman G and Luttik R. Reference values for comparing the acute toxicity of pesticides to birds. Reviews of Environmental Contamination and Toxicology, (2000) In Press.

Mineau, P.; Collins, B. T.; Baril, A. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Regul. Toxicol. Pharmacol. 24(1, Pt. 1), 24-29 (1996).

Moody, Richard P. Dermal penetration.:in Ecobichon, Donald J. Occup. Hazards Pestic. Exposure: Sampling, Monit., Meas. (1999), 81-105. Publisher: Taylor & Francis, Philadelphia, Pa

Novartis Crop Protection. 1997. An ecological risk assessment of diazinon in the Sacramento and San Joaquin River basins. Technical Report: 11/97. Novartis Crop Protection, Inc., Greensboro, NC.

Power, M.; McCarty, L. S. Fallacies in ecological risk assessment practices Environ. Sci. Technol., 31(8), 370 A-375 A (1997).

Roth, Charles M.; Yarmush, Martin L. Nucleic acid biotechnology. Ann. Rev. Biomed. Eng. (1999), 1 265-297.

Sample, B. E.; Arenal, C. A. Allometric models for interspecies extrapolation of wildlife toxicity data. Bull. Environ. Contam. Toxicol. 62(6), 653-663. (1999).

Shapira, Michael; Seidman, Shlomo; Livni, Nadav; Soreq, Hermona.. In vivo and in vitro resistance to multiple anticholinesterases in Xenopus laevis tadpoles Toxicol. Lett. 102-103 205-209(1998).

Solomon, K.R. and J.M. Giddings. 2000. Understanding probabilistic risk assessment. SETAC Globe: Learned Discourse,Vol. 1 (1): 34-35.

Walker, Colin H.. Biomarker strategies to evaluate the environmental effects of chemicals Environ. Health Perspect. Suppl. (1998), 106(2), 613-620.

Wecker, Lynn; Dettbarn, Wolf D. Paraoxon-induced myopathy: muscle specificity and acetylcholine involvement. Exp. Neurol. (1976), 51(2), 281-91.

Wilson, B W, Cholinesterases in "The Clinical Chemistry of Laboratory Animals", 2nd edition, Loeb W F and Quimby FW, Editors, Taylor and Francis pp 430-440 (1999).

Winzeler, Elizabeth A.; Schena, Mark; Davis, Ronald W. Fluorescence-based expression monitoring using microarrays**.** Methods Enzymol. (1999), 306 (Expression of Recombinant Genes in Eukaryotic Systems), 3-18, 1 Plate.

SAP Report No. 00-02B,
August 2, 2000

REPORT:

FIFRA Scientific Advisory Panel Meeting, April 7, 2000, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Insect Repellent Product Performance Testing Guideline Evaluation

Mr. Larry Dorsey
Designated Federal Official
FIFRA/Scientific Advisory Panel
Date:_____

Mary Anna Thrall, DVM
FIFRA SAP Session Chair
FIFRA/Scientific Advisory Panel
Date:_____

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
April 7, 2000

SESSION II - Insect Repellent Product Performance Testing Guideline Evaluation

PARTICIPANTS

FIFRA Scientific Advisory Panel Session Chair
Mary Anna Thrall, D.V.M., Professor, Department of Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO
FIFRA Scientific Advisory Panel
Charles Capen, D.V.M, Professor and Chairman, Department of Veterinary Biosciences, The Ohio State University, Columbus, OH
Herbert Needleman, M.D., Professor of Psychiatry and Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA

FQPA Science Review Board Members
Stephen B. Berte, Ph.D., Lieutenant Colonel, U.S. Army, Contingency Liaison Officer, Armed Forces Pest Management Board, Silver Spring, MD
Jonathan Day, Ph.D., Florida Medical Entomology Laboratory, IFAS, University of Florida, Vero Beach, FL
CDR Russell Enscore, Environmental Health Specialist, Centers for Disease Control, Rampart Road–Foothills Campus, Fort Collins, CO
Durland Fish, Ph.D., Yale University, Department of Epidemiology and Public Health, New Haven, CT
Eugene Gerberg, Ph.D., Professor, University of Florida, Department of Entomology Gainesville, FL
Raj Gupta, Ph.D., US Army Medical Research, Ft. Dietrick, MD
Robert J. Novak, Ph.D., Illinois Natural History Survey, Champaign, IL
Kenneth M. Portier, Ph.D., Associate Professor, Department of Statistics, Institute of Food and Agricultural Sciences and Natural Resources, University of Florida, Gainesville, FL
Andrew Spielman, Ph.D., Harvard School of Public Health, Boston, MA

Designated Federal Official
Mr. Larry Dorsey, FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, DC

PUBLIC COMMENTERS

Oral statements were made by:
Mr. Marvin Bertsch, Mr. Nick Spero, and Robin Todd, Ph.D., on behalf of Insect Control and Research, Inc.
Scott Carroll, Ph.D. and Jenella Lloye, Ph.D., on behalf of Carroll-Loye Biological Research
Stephen Gettings, Ph.D., on behalf of Avon Products, Inc., Product Safety and Integrity
Peter Gray, Esq., McKenna & Cuneo, L.L.P., on behalf of the Chemical Specialties Manufacturers Association
James Hudson, Ph.D. on behalf of Pest Management Regulatory Agency, Health Canada
Daniel Lawson, Ph.D., DEET Issues Task Force, on behalf of the Chemical Specialties Manufacturers Association
Ms. Julie Spagnoli, on behalf of Bayer Corporation

Written statements were received from:
Mr. Marvin Bertsch, Mr. Nick Spero, and Robin Todd, Ph.D., on behalf of Insect Control and Research, Inc.
Scott Carroll, Ph.D. and Jenella Lloye, Ph.D., on behalf of Carroll-Loye Biological Research
Stephen Gettings, Ph.D., on behalf of Avon Products, Inc., Product Safety and Integrity
Daniel Lawson, Ph.D., DEET Issues Task Force, on behalf of the Chemical Specialties Manufacturers Association

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 pertaining to Insect Repellent Product Performance Testing Guideline Evaluation. Advance notice of the meeting was published in the *Federal Register* on March 16, 2000. The review was conducted in an open Panel meeting held in Arlington, Virginia, on April 7, 2000. The meeting was chaired by Mary Anna Thrall, D.V.M. Mr. Larry Dorsey served as the Designated Federal Official.

Inconsistencies have developed in product performance testing and labeling of insect repellents. In order to minimize this variance, EPA has developed draft product performance testing guidelines and appropriate label language. This guideline recommends specific methods for conducting product performance testing of insect repellents. As a guideline, it does not impose mandatory requirements. It does, however, reflect the Agency's considered recommendations for minimum steps necessary to develop reliable data on repellent product performance. In addition, the product performance testing guidelines are intended to supersede EPA, Pesticide Assessment Guidelines, Subdivision G: 95-9, ``Treatments to control pests of humans and pests" and 95-10, ``Mosquito, black fly, nonbiting midge, and biting midge."

A performance standard represents the minimum level of product performance which would normally be acceptable for protecting public health, when required, or for economic control of a pest or pest combination at a specific site. These guidelines are concerned with product performance testing for evaluation of pesticides used to repel mosquitos, biting flies, fleas, chiggers and ticks from human skin and outdoor premises. EPA intends to use the data from guideline studies to help determine the adequacy of the labeling of insect repellant products. The label language proposed by the Agency is intended to standardize and improve the information provided to the consumer. The Agency sought the Panel's advice on the adequacy of the proposed testing guidelines and protocols for human insect repellants. Ms. Robyn Rose (EPA, Office of Pesticide Programs), Mr. Kevin Sweeney (EPA, Office of Pesticide Programs), and Russell S. Jones, Ph.D. (EPA, Office of Pesticide Programs) provided an introduction and summary of insect repellent product performance testing guidelines. Mr. Larry Dorsey served as the Designated Federal Official.

CHARGE

The specific issues to be addressed by the Panel are keyed to the background document, "OPPTS 810.3700; Insect Repellents for Human Skin and Outdoor Premises" memorandum dated March 16, 2000, and are presented as follows:

FIRST BITE vs. FIRST CONFIRMED BITE vs. 95% REDUCTION IN BITES

1. First Bite (FB) vs. First Confirmed Bite (FCB): Historically, the Agency has used the First Confirmed Bite (FCB) test to assess the effectiveness of human insect repellents. However, the Agency is concerned that the FCB method will result in the loss of valuable information. The FCB method does not appear to have been developed using a statistically valid approach. For this reason and because some insect bites may be disregarded when all bites should be counted, the Agency does not currently approve of the FCB method. The Agency recommends use of the First Bite (FB) method or a 95% reduction in bites, because all bites are counted and the method provides a more "real-world" assessment of insect repellent efficacy.

Is the Panel aware of any scientifically valid justification for using the FCB method, or, conversely with using the FB or 95% reduction in bites methods. Should we use 95% and a first bite test or choose just one of these as the standard - why or why not?

GENERAL CONSIDERATIONS FOR ALL TESTING

2. If a product effectively repels a particular pest based upon the time to first bite, the Agency is considering allowing a claim of protection against potential disease vectors. For example: "May repel deer ticks which carry lyme disease."

What degree of protection is necessary to warrant allowing claims of protection from specific diseases? What rationale can the Agency use to demonstrate a high enough level of

efficacy to claim protection against potential disease vectors? What suggestions if any does the Panel have for changes to these protocols that would allow a claim for protection against potential disease vectors? Can you suggest a way to account for differences in level of repellency for different products?

3. The Agency is recommending five treated test subjects for a label claim of less than five hours of repellency and ten treated test subjects for a label claim of five or more hours of repellency.

The Agency considered the publications by Rutledge and Gupta (1999) as a resource in the development of recommendations for the numbers of replications to be used in field tests of insect repellents (Appendix I). Although the Agency believes that the data are scientifically sound, a direct and literal use of these data may not be practical (either economically or logistically) for all registrants. However, after review of Rutledge and Gupta (1999), the Agency realized that more test subjects may be necessary to test repellents with longer durations of repellency.

What number of test subjects would provide statistically-valid results? If more test subjects then currently recommended by EPA are appropriate, would it then be feasible for Registrants to conduct the test? If the number of test subjects should be different for repellents with shorter claims of duration of repellency, how many test subjects should repellents with longer claims include?

4. How should exposure testing be designed to take into account that some test organisms (e.g., mosquitoes) only bite during specific times in a day which may exceed the duration of repellency. For example, would it be acceptable to apply repellent to test subjects at varying number of hours before exposure (e.g., 1,2,4,8, and 12 hours) and then expose all subjects at once? Why, or why not? For this method, how many times should each test subject be exposed? Can you recommend an alternative way to address this problem that might be better?

5. Are the application rates proposed in "OPPTS 810.3700; Insect repellents for human skin and outdoor premises" acceptable for a scientifically sound study? If not, how should application rates be derived? Should an application rate be recommended in these protocols or left to the discretion of the registrant? If a repellent is applied as a thick layer, how will it affect the results of the efficacy test?

MOSQUITO AND STABLE FLY LABORATORY TESTS

6. How valuable are cage studies in assessing the efficacy of a repellent? If the Agency decides to require submission of the cage studies, are there better ways to perform the studies than the Agency-recommended protocols? If so, what are they? Are there advantages to the Klun and Debboun (2000) study that might justify including it as an alternative method (Appendix 2)? If so, what are they?

MOSQUITO, BLACKFLY CERATOPOGONID, SANDFLY, TABANID, AND STABLE FLY FIELD TESTS

7. What biting pressures are appropriate, e.g., five bites in ten minutes for Ceratopogonids and one bite in five minutes for Tabanids? How should biting pressure be determined, e.g., should lands be considered as well as probes and/or bites? If landing rate data collection can be justified for laboratory and/or field studies, what rates would be acceptable?

CANDLES, COILS, AND VAPORIZING MATS

8. The agency has proposed a 50% reduction in bites for a label claim that the repellent may aid in reducing bites and a 95% reduction in bites for a label claim that the product repels, e.g., mosquitoes. What level of reduction in bites is acceptable to show efficacy for candles, coils, and vaporizing mats?

FLEAS

9. What laboratory tests will provide adequate data to determine flea repellency? Of those, including the USDA test found in Appendix III, are any better than the Agency-proposed tests? How many lands should be required within three or five minutes to verify biting pressure (e.g., the Agency proposed ten)?

TICKS AND CHIGGER MITES

10. Due to the high incidence of Lyme disease in the U.S., EPA did not recommend deer tick field tests using human subjects. How adequate are the proposed laboratory tests in determining deer tick repellency? Evaluate the tick and chigger tests found in Appendix III (Smith 1955) and IV. Should these protocols be considered in lieu of or in addition to the Agency proposal?

DETAILED RESPONSE TO THE CHARGE

FIRST BITE vs. FIRST CONFIRMED BITE vs. 95% REDUCTION IN BITES

1. First Bite (FB) vs. First Confirmed Bite (FCB): Historically, the Agency has used the First Confirmed Bite (FCB) test to assess the effectiveness of human insect repellents. However, the Agency is concerned that the FCB method will result in the loss of valuable information. The FCB method does not appear to have been developed using a statistically valid approach. For this reason and because some insect bites may be disregarded when all bites should be counted, the Agency does not currently approve of the FCB method. The Agency recommends use of the First Bite (FB) method or a 95% reduction in bites, because all bites are counted and the method provides a more "real-world" assessment of insect repellent efficacy.

Is the Panel aware of any scientifically valid justification for using the FCB method, or, conversely with using the FB or 95% reduction in bites methods. Should we use 95% and a first bite test or choose just one of these as the standard - why or why not?

The consensus of the Panel was that the 95% reduction in biting should be the principal standard for testing repellents. The Panel's decision is based on the application of good science in the experimental design (including the use of an untreated control) and subsequent data analysis. In addition, several Panel members commented that the 95% reduction method provides a stronger basis for the data to be statistically analyzed. The Panel also agreed that the 95% reduction in bite method is more easily understood than either the First Confirmed Bite or First Bite. The first bite methods could be utilized to establish the time period of complete protection for a repellent. While the specific time for complete protection was discussed later in response to Question 3, a 2 hour minimum was suggested. It was also suggested that the Agency adopt a standard scientifically-based testing protocol with subsequent review and comment by the FIFRA SAP. One possible design could be a latin-square design. Development of such a protocol would dictate the standards for testing, thus helping to alleviate GLP concerns. In any case, reducing vector borne diseases should not be used as a rationale in the development of testing protocols.

GENERAL CONSIDERATIONS FOR ALL TESTING

2. If a product effectively repels a particular pest based upon the time to first bite, the Agency is considering allowing a claim of protection against potential disease vectors. For example: "May repel deer ticks which carry lyme disease."

What degree of protection is necessary to warrant allowing claims of protection from specific diseases? What rationale can the Agency use to demonstrate a high enough level of efficacy to claim protection against potential disease vectors? What suggestions if any does the Panel have for changes to these protocols that would allow a claim for protection against potential disease vectors? Can you suggest a way to account for differences in level of repellency for different products?

The consensus of the Panel was that no claim should be made regarding protection against arthropod-borne pathogens. There are several points presented by the Panel to support this position. First, most arthropods that interact with humans and/or animals are not capable of transmitting pathogens. In addition, in those instances where a potential disease vector exists, the Panel cautioned against a claim of repellency for the products. Gupta & Rutledge (1994) reported that the use of repellents to reduce human vector contact and reduce the transmission of mosquito-borne diseases has not been scientifically proven. Second, individual factors such as proper application, individual variability and susceptibility, and environmental factors (temperature, humidity, perspiration production, rain, clothing presence), also affect the degree of protection afforded by the repellent. Therefore, in order for the Agency to rely on the best scientific data for claims of insect repellency, the use of repellents for reducing arthropod-borne pathogens must be determined.

3. The Agency is recommending five treated test subjects for a label claim of less than five hours of repellency and ten treated test subjects for a label claim of five or more hours of repellency.

The Agency considered the publications by Rutledge and Gupta (1999) as a resource in the development of recommendations for the numbers of replications to be used in field tests of insect repellents (Appendix I). Although the Agency believes that the data are scientifically sound, a direct and literal use of these data may not be practical (either economically or logistically) for all registrants. However, after review of Rutledge and Gupta (1999), the Agency realized that more test subjects may be necessary to test repellents with longer durations of repellency.

What number of test subjects would provide statistically-valid results? If more test subjects then currently recommended by EPA are appropriate, would it then be feasible for Registrants to conduct the test? If the number of test subjects should be different for repellents with shorter claims of duration of repellency, how many test subjects should repellents with longer claims include?

The Panel suggests that primary emphasis regarding sample size (human test subjects) should be based on the scientific experimental design and not on formula driven guidelines. It was pointed out that there are inherent flaws in the determination of sample size in Gupta and Rutledge (1979). For example, according to Ruthledge and Gupta, for five individuals the confidence of protection is 97.5% confidence protection for 1 hour but at 2 hours it is only about 50%. In Table 4 of the Agency's Background Document (No. Subjects, Protection Periods 1-8 hours, Confidence limit 99 and 95 %) the best possible results ($P < 0.01$ with $D = 0.5$ h) for a product claiming 1 hour of protection would require 15 test subjects whereas one claiming 8 hours requires 280. This would not be feasible or practical.

In most experimental designs for the evaluation of insect repellents, gaining an adequate number of replications of the product(s) is typically stressed over using a large number of subjects. The principal objective is to ensure that tests are replicated a sufficient number of times in order to strengthen the power of associated statistical tests. Thus, for example, if four repellent concentrations and a control are tested using five individuals, the whole assessment could be repeated 5 times. In this way, the assignment of treatments to individuals over replicate assessment "rounds" would be such that each individual would eventually be evaluated on each treatment (a "round robin" or Latin Square design). This is just one of a set of equally acceptable study designs. The Panel suggested that the Agency, rather than proscribing evaluation protocols, consider a solution similar to that used in the National Institutes of Health to evaluate their assessment studies. In particular, the Panel suggested that convening an expert panel, such the FIFRA SAP, to periodically evaluate, comment on and recommend changes to industry-proposed study protocols might be the most effective way of handling testing protocol specification. Acceptable design protocols could be published and most new assessments would be performed using these protocols.

4. How should exposure testing be designed to take into account that some test organisms (e.g., mosquitoes) only bite during specific times in a day which may exceed the duration of repellency. For example, would it be acceptable to apply repellent to test subjects at varying number of hours before exposure (e.g., 1, 2, 4, 8, and 12 hours) and then expose all subjects at once? Why, or why not? For this method, how many times should each test subject be exposed? Can you recommend an alternative way to address this problem that might be better?

Since insects/arthropods seek hosts at different times of the day, it is essential that the field testing of repellents occur at those times. In addition, repellents should be applied at different times to establish efficacy. Exposing human subjects to continuous biting activity as proposed is unnecessary. It is feasible to apply repellents to human skin surfaces of all volunteers at one time and then to expose all volunteers together to coincide with arthropod activity periods. The test subjects should remain in a field environment to simulate climatic conditions that properly test efficacy under an actual use scenario. Numerous studies have indicated that repellency can be influenced by changes in temperature, humidity, rate of perspiration, physical activity, and abrasion with clothing. Each test subject should be rotated through all the treatment regimens, including the untreated control, to reduce inter-personal effects due to differential attractiveness of individuals to insects and variability of individual effectiveness of repellents. If insufficient statistical power is achieved with this approach, additional subjects could be used to increase the per test number of individuals exposed at each dose.

5. Are the application rates proposed in "OPPTS 810.3700; Insect repellents for human skin and outdoor premises" acceptable for a scientifically sound study? If not, how should application rates be derived? Should an application rate be recommended in these protocols or left to the discretion of the registrant? If a repellent is applied as a thick layer, how will it affect the results of the efficacy test?

The Panel believes that the amount of repellent to be applied to the skin could be determined by the registrant for several reasons. The proposed guidelines (OPPTS 810,3700) specifies that the applied product amount should be determined by weight. The Panel disagreed with this approach and specifically saw potential problems when dealing with application rates of aerosols. Therefore, since most repellents are liquids, creams or aerosols, the application rates should be in milliliters (or in seconds of spray time for aerosol). In addition, the test area for application of 600 cm^2 is too large an area for many arms. A test area of 250-300 cm^2 is more than adequate.

The amount of the repellent to be tested should be determined by conducting statistically valid studies that demonstrate the quantity of a given physical formulation consumers are likely to apply. Based on public comments from an industry representative at the meeting, it is apparent some of these data already exist in the cosmetic industry. If such data are not available, repellent

manufacturers should conduct such studies to provide such data to the Agency. The dose rate per unit could than be established through pre-field tests using cage tests. The rationale for this is that there are, and in the future, will be numerous new products that do not fit the synthetic chemical repellent mode of action. We are already seeing this with the increased number of natural repellents and many new products that have multiple purposes, i.e. sun-screen, moisturizers etc. This would certainly play a major role in determining the application amount. Field efficacy data could then be used by the Agency for registration.

MOSQUITO AND STABLE FLY LABORATORY TESTS

6. How valuable are cage studies in assessing the efficacy of a repellent? If the Agency decides to require submission of the cage studies, are there better ways to perform the studies than the Agency-recommended protocols? If so, what are they? Are there advantages to the Klun and Debboun (2000) study that might justify including it as an alternative method (Appendix 2)? If so, what are they?

The Panel strongly recommends that only field studies be used to establish efficacy and subsequent registration. Cage studies are not a valid substitute for repellent field studies but they can be used to compare products. Cage tests should be used only as a screening device and should not be submitted in support of a registration. They could, however, be used by the manufacturer to screen possible repellents, developing formulations, and determining a range of application rates.

The Klun & Debboun device may be an alternative to the device specified in the ASTM Standard for laboratory studies of mosquitoes (ASTM 951-94). However, it is a screening tool that was never intended as a substitute for mosquito field studies. If a test cage with an enclosed area, such as Klun & Debboun, does not provide for free flow of repellent vapors from the surface and eventual dissipation of repellent vapors into immediate environment, it is probable that some repellents may have erroneously indicated higher repellency. Any laboratory test cage selected for product testing should take the vaporous state of repellents into account before being recommended for use.

MOSQUITO, BLACKFLY CERATOPOGONID, SANDFLY, TABANID, AND STABLE FLY FIELD TESTS

7. What biting pressures are appropriate, e.g., five bites in ten minutes for Ceratopogonids and one bite in five minutes for Tabanids? How should biting pressure be determined, e.g., should lands be considered as well as probes and/or bites? If landing rate data collection can be justified for laboratory and/or field studies, what rates would be acceptable?

The recommendation of the Panel for biting pressures appropriate for testing are based primarily on what the general public perceives as a nuisance problem. The Panel would recommend the following biting rates for field-testing: mosquitoes 1 bite per minute;

ceratopogonids at 1 bite per 5 minutes; tabanids at 1 bite per 5 minutes. Since little information is available in the literature, the experience of members of the Panel coupled with a publication by Morris and Clanton (1988) titled *Quantification of a nuisance mosquito problem in Florida* were used as guidelines. It is important to remember that repellents are typically used for nuisance problems rather than for disease prevention. Therefore, it follows that the guidelines used by the Agency regarding biting pressure reflect conditions that impact the general public and not military or public health personnel.

CANDLES, COILS, AND VAPORIZING MATS

8. The Agency has proposed a 50% reduction in bites for a label claim that the repellent may aid in reducing bites and a 95% reduction in bites for a label claim that the product repels, e.g., mosquitoes. What level of reduction in bites is acceptable to show efficacy for candles, coils, and vaporizing mats?

A 50% repellency of mosquitoes and other arthropods is not appropriate. If candles, coils, vaporizing mats or other such products are to be useful, they should provide at least 95% repellency. FIFRA SAP member Robert Novak will submit to EPA a manuscript that is in press in the Journal of the American Mosquito Control Association, where coils, candles, plants etc. are tested under field conditions.

FLEAS

9. What laboratory tests will provide adequate data to determine flea repellency? Of those, including the USDA test found in Appendix III, are any better than the Agency-proposed tests? How many lands should be required within three or five minutes to verify biting pressure (e.g., the Agency proposed ten)?

Even though the proposed tests by the Agency to evaluate flea repellency is adequate, the Panel questions whether a flea repellent for humans is necessary in North America. In any event, the test proposed by the Agency is adequate to evaluate repellency. The Agency should recognize that alternative test protocols have been developed and published by laboratories throughout the world, many of which are perfectly adequate to determine flea repellency. The USDA test is somewhat lacking in detail and, as such, comparison of data from one test to another may be called into question. The Agency should have the flexibility to evaluate and allow alternative testing methods.

Flea bites can be painful, allergenic, and annoying. When first disturbed, fleas will jump and inadvertently land on hosts. However, when in the blood feeding mode, they will land and walk on a host in search of a feeding site. It is the Panel's recommendation that one landing and probable probe per minute should be the standard used to verify biting and lack of repellency.

TICKS AND CHIGGER MITES

10. Due to the high incidence of Lyme disease in the U.S., EPA did not recommend deer tick field tests using human subjects. How adequate are the proposed laboratory tests in determining deer tick repellency? Evaluate the tick and chigger tests found in Appendix III (Smith 1955) and IV. Should these protocols be considered in lieu of in addition to the Agency proposal?

The Panel does not agree with the Agency's recommendation against deer tick field tests on human subjects because of disease risk. It is precisely because of high risk of disease transmission from deer tick bites in Lyme disease endemic areas that field-testing with human subjects should be required. There are several reasons and options available to do field testing which minimize the risks to subjects. First, the black-legged tick *Ixodes scapularis* does not carry the Lyme disease spirochete throughout its range and prevalence of infection is insignificant (<1%) over most of its range. This is also true for the western black-legged tick *Ixodes pacificus* where prevalence rates in excess of 1% are rare. Under these circumstances, it seems unreasonable to identify the black-legged tick or the western black-legged tick, as species having exceptionally high risk for disease transmission to human subjects. In areas with high infection prevalence, effective measures can be taken to minimize infection and prevent disease among subjects. These include careful inspection for attached ticks following exposure. Transmission of the Lyme disease spirochete does not occur within the first 48 hours of attachment. Additionally, subjects may be offered a vaccine or prophylactic antibiotics dose prior to exposure to prevent infection. Alternatively, subjects can be tested for antibodies to tick borne infections before and two weeks following exposure to detect asymptomatic infection. Subjects should be briefed on the symptoms of tick-borne infections and seek medical treatment from any unusual symptoms following exposure to ticks. These recommendations would be appropriate for all field tests involving ticks, regardless of the species, location or perceived risk of infection.

The Panel recommends that the Agency review the scientific literature to evaluate a laboratory test method for ticks and chiggers that do not involve use of humans directly. Such studies could include a method used by Buescher et al 1984.

The Panel questions why only tick repellents are permitted to claim efficacy against certain tick species. There is no evidence to support the notion that response variability among tick species is greater than that of mosquitoes. Since nymphal stage dog ticks do not feed upon humans, the Panel concludes there is no justification for requiring testing different stages based upon disease potential. Adult "deer" ticks are equally capable of transmitting ehrlichiosis as the nymphs.

More research is needed to develop an improved field test for mites and chiggers. Laboratory tests do not seem sufficient to determine efficacy of tick or chigger repellents. The alternative methods supplied in the appendix of the Agency's background documents do not seem to be much of an improvement. There is a serious need for research on tick and mite behavioral biology that would provide critical information pertinent to the issue of repellent testing.

ADDITIONAL COMMENTS

The Panel also provided additional comments as provided below.

(1) The Panel strongly recommends that the Agency have a mathematical statistician spend some time with the problem of experimental design. There are statistical models/tools that would clarify many of the sample size/statistical power issues.
(2) The Agency should not require GLP standards for field trials. First, GLPs were designed for laboratory studies. Also, if the Agency requires a scientifically based experimental design, the standards for GLP for lab studies would be incorporated for field studies. The GLP standards that are used for field studies do not fit nor add anything to the quality of the field test except additional costs.
(3) Several Panel members suggested that both male and female test subjects be utilized in field tests to evaluate gender-related efficacy differences.