UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

MEMORANDUM

DATE: April 30, 2012

SUBJECT: Transmittal of the Meeting Minutes of the FIFRA SAP Meeting Held January 31 to February 2, 2012 on the Scientific Issues Associated with "Common Effects Assessment Methodology Developed in the Office of Pesticide Programs and Office of Water"

FROM: Sharlene Matten, Ph.D.
Designated Federal
FIFRA SAP Staff
Office of Science Coordination and Policy

THRU: Laura Bailey
Executive Secretary, FIFRA SAP
Office of Science Coordination and Policy

Frank Sanders
Director
Office of Science Coordination and Policy

TO: Steven Bradbury, Ph.D.
Director
Office of Pesticide Programs

Please find attached to this memorandum the meeting minutes of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) open meeting held in Arlington, Virginia on January 31-February 2, 2012. This report addresses a set of scientific issues associated with "Common Effects Assessment Methodology Developed in the Office of Pesticide Programs and Office of Water."

Attachment

cc:
Jim Jones
Vicki Dellarco
William Jordan
Margie Fehrenbach
Donald Brady
Joan Harrigan-Farrelly
Jack Housenger
Richard Keigwin, Jr.
Keith Matthews
Robert McNally
Lois Rossi
Karen Whitby
Oscar Morales
Douglas Parsons
Enesta Jones
Vanessa Vu
OPP Regulatory Docket
Elizabeth Behl
Mark Corbin
Kristina Garber
Joe Beaman
Christine Russom
Matthew Etterson
Russell Erickson
Glen Thursby
David Mount
Dale Hoff

FIFRA SAP Members
Kenneth Portier, Ph.D. (FIFRA SAP Chair)
Janice Chambers, Ph.D., DABT, ATS
(Session Chair)
Stephen Klaine, Ph.D.
Martha Sandy, Ph.D.
Daniel Schlenk, Ph.D.

FQPA Science Review Board Members
Brian S. Anderson, M.A.
Bryan W. Brooks, Ph.D.
G. Allen Burton, Jr., Ph.D.
Peter M. Chapman, Ph.D., R.P.Bio.
Mark T. D. Cronin, Ph.D.
Nancy D. Denslow, Ph.D.
Cheryl A. Murphy, Ph.D.
James T. Oris, Ph.D.
Geoffrey I. Scott, Ph.D.

SAP Minutes No. 2012-02

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

Comparative Effects Methodology Developed by the Office of Pesticide Programs and the Office of Water

January 31 to February 2, 2012
FIFRA Scientific Advisory Panel Meeting
Held at
One Potomac Yard
Arlington, Virginia

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). The meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (EPA or Agency). The content of the meeting minutes does not represent information approved by the Agency. The meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of these meeting minutes 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 is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA 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 Sharlene R. Matten, Ph.D., SAP Designated Federal Official, via e-mail at matten.sharlene@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented in public comment. This document addresses the information provided and presented by the EPA within the structure of the charge.

TABLE OF CONTENTS

SAP Minutes No. 2012-02

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

Comparative Effects Methodology Developed by the Office of Pesticide Programs and the Office of Water

January 31 to February 2, 2012
FIFRA Scientific Advisory Panel Meeting
Held at
One Potomac Yard
Arlington, Virginia

Janice E. Chambers, Ph.D., DABT, ATS
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel

Janice E. Chambers

Date: April 27, 2012

Sharlene R. Matten, Ph.D.
Designated Federal Official
FIFRA Scientific Advisory Panel
Staff

S.R. Matten

Date: April 27, 2012

Panel Member List for the Meeting of the Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) to Consider and Review Comparative Effects Methodology Developed by the Office of Pesticide Programs and the Office of Water

January 31-February 2, 2012

EPA-HQ-OPP-2011-0898 **OPP Docket Tel: 703-305-5805**

FIFRA SAP Chair

Kenneth M. Portier, Ph.D.
Program Director, Statistics
American Cancer Society
National Home Office
Atlanta, Georgia

FIFRA SAP Session Chair

Janice E. Chambers, Ph.D., DABT, ATS
William L. Giles Distinguished Professor
Director, Center for Environmental Health Sciences
College of Veterinary Medicine
Mississippi State University
Mississippi State, Mississippi

Designated Federal Official

Sharlene R. Matten, Ph.D.
US Environmental Protection Agency
Office of Science Coordination & Policy
FIFRA Scientific Advisory Panel
EPA East Building, MC 7201M
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Tel: 202-564-0130, Fax: 202-564-8382, matten.sharlene@epa.gov

FIFRA Scientific Advisory Panel Members

Stephen J. Klaine, Ph.D.
Professor and Director
Clemson University Institute of Environmental Toxicology
Department of Biological Sciences
Pendleton, South Carolina

Martha S. Sandy, Ph.D.
Senior Toxicologist and Chief
Cancer Toxicology and Epidemiology Section
Reproductive and Cancer Hazard Assessment Branch
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency
Oakland, California

Daniel Schlenk, Ph.D.
Professor of Aquatic Ecotoxicology and Environmental Toxicology
Department of Environmental Sciences
University of California, Riverside
Riverside, California

FQPA Science Review Board Members

Brian S. Anderson, M.A.
Research Specialist
Marine Pollution Studies Laboratory
University of California, Davis
Granite Canyon, California

Bryan W. Brooks, Ph.D.
Professor and Director
Environmental Health Science Program
Environmental Science and Biomedical Studies
Center for Reservoir and Aquatic Systems Research
Baylor University
Waco, Texas

G. Allen Burton, Jr., Ph.D.
Professor and Director
Cooperative Institute for Limnology and Ecosystems Research
Professor, School of Natural Resources and Environment &
Department of Earth & Environmental Sciences
University of Michigan
Ann Arbor, Michigan

Peter M. Chapman, Ph.D., R.P.Bio.
Principal and Senior Environmental Scientist
Golder Associates Ltd.
Burnaby, British Columbia
Canada

Mark T. D. Cronin, Ph.D.
Professor of Predictive Toxicology
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Liverpool, England

Nancy D. Denslow, Ph.D.
Professor
Department of Physiological Sciences
University of Florida
Gainesville, Florida

Cheryl A. Murphy, Ph.D.
Assistant Professor
Department of Fisheries and Wildlife
Michigan State University
East Lansing, Michigan

James T. Oris, Ph.D.
Interim Associate Provost for Research and Scholarship &
Dean of the Graduate School
Professor of Zoology
Miami University
Oxford, Ohio

Geoffrey I. Scott, Ph.D.
Director
Charleston Center for Coastal Environmental Health and Biomolecular Research
National Oceanic & Atmospheric Administration, National Ocean Services
US Department of Commerce
Charleston, South Carolina

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed its report of the SAP meeting regarding scientific issues associated with **"Comparative Effects Methodology Developed by the Office of Pesticide Programs and the Office of Water."** Advance notice of the SAP meeting was published in the *Federal Register* on **November 16, 2011**. The review was conducted in an open Panel meeting on **January 31-February 2, 2012** at One Potomac Yard, Arlington, Virginia. Materials for this meeting are available in the Office of Pesticide Programs public docket or via www.regulations.gov, **Docket No. EPA-HQ-OPP-2011-0898**. Janice Chambers, Ph.D., DABT, ATS chaired the meeting and Sharlene Matten, Ph.D., served as the Designated Federal Official. Donald Brady, Ph.D., Director, Environmental Fate and Effects Division, Office of Pesticide Programs (OPP), United States Environmental Protection Agency (EPA) and Elizabeth Behl, M.S., Director, Health and Ecological Criteria Division, Office of Science and Technology, Office of Water (OW), EPA, provided opening remarks at the meeting. Agency presentations of technical background materials were provided by:

1) Kristina Garber, M.S., Environmental Fate and Effects Division, OPP;
2) Joe Beaman, M.S., Health and Ecological Criteria Division, Office of Science and Technology, OW;
3) Christine Russom, Mid-Continent Ecology Division, Office of Research and Development;
4) Matthew Etterson, Ph.D., Mid-Continent Ecology Division, Office of Research and Development;
5) Russell Erickson, Ph.D., Mid-Continent Ecology Division, Office of Research Development; and,
6) Glen Thursby, Ph.D., Atlantic Ecology Division, Office of Research and Development.

Additional technical clarifications were provided via teleconference by David Mount, Ph.D. and Dale Hoff, Ph.D. from the Mid-Continent Ecology Division, Office of Research and Development, EPA.

The mission of the EPA is to protect human health and the environment. Consistent with the EPA's mission, OPP and OW are both responsible for scientifically evaluating potential effects of chemicals on aquatic life. This is accomplished by OPP through risk assessment and OW through aquatic life criteria (ALC) development. For pesticides, data are available to inform deterministic, and in some cases, probabilistic risk assessments used in pesticide registration by OPP; however, in many cases the databases used by OPP for risk assessment contain fewer studies than OW uses to derive ALC according to OW's 1985 Guidelines[1]. Currently, OW has 47 established national recommended ALC[2]. Of these, 16 are pesticides that are currently registered for use in the US. Meanwhile, OPP has registered over 1000 active ingredients that may be used in the US. Surface water monitoring reports (e.g., by the United States Geological Survey's

[1] Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses (USEPA 1985), referred to as the 1985 Guidelines

[2]Current National Recommended Water Quality Criteria are available online at: http://www.epa.gov/waterscience/criteria/wqctable/index.html

National Water-Quality Assessment Program) for certain chemicals, including many pesticides, have prompted stakeholders to raise questions as to the possible risks these chemicals pose to aquatic communities. No ALC have been established for many of these chemicals, which impedes the ability of stakeholders to interpret the potential effects on aquatic communities where the chemicals are detected. OPP's Aquatic Life Benchmarks, which are based on toxicity values used in OPP's most recent risk assessments developed as part of the decision-making process for pesticide registration, allow for comparison of measured concentrations to acute and chronic endpoints for fish, invertebrates and plants; however, they are not designed to represent community level benchmarks.

The EPA's current effort involves exploration of several approaches that may be considered with varying amounts of data for OPP and OW to characterize the effects of chemicals on aquatic organisms. Emphasis is placed on situations where data that OPP would regard as sufficient are available to conduct risk assessments, but OW would not have a data set that, according to the 1985 Guidelines, would support derivation of ALC. The EPA sought advice from the SAP on the following tools and methods analyzed in the white paper and appendices.

1) Predictive tools, i.e., (Quantitative) Structure-Activity Relationships ((Q)SARs) and Interspecies Correlation Estimation (ICE) used to estimate the acute toxicity of aquatic animals exposed to pesticides.
2) Potential applications of species sensitivity distributions (SSDs) and extrapolation factors (EFs) for characterizing the effects of chemicals within taxa and communities of aquatic animals and plants.
3) The uses of acute-chronic ratios (ACRs) to estimate chronic effect thresholds for aquatic animals lacking chronic toxicity data.
4) Use of these approaches to reliably predict the fifth percentiles of sensitivity distributions of toxicity data (termed the “HC5”), which are intended to be representative of more sensitive species in aquatic communities of animals and plants and serve as the assessment endpoint used by OW for derivation of ALC.
5) Potential influences of pesticidal modes of action (MOA), e.g., acetylcholinesterase (AChE) inhibitors, on the utility of these methods.

PUBLIC COMMENTERS

Oral statements were provided on behalf of CropLife America by:

1) Nicholas Poleticka, Ph.D., Dow AgroSciences, LLC;
2) Dwayne Moore, Ph.D., Intrinsik Environmental Sciences, Inc.; and,
3) Jeffrey Giddings, Ph.D., Compliance Services International (CSI).

Written statements were provided by:

Michael Leggett, Ph.D., CropLife America

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Charge Question 1: ((Quantitative) Structure-Activity Relationships ((Q)SARs) Tools to Predict Acute Toxicity

Appendix A of the white paper includes a review of the utility of ECOSAR, TEST and the OECD toolbox for quantifying sensitive acute toxicity values for aquatic animals exposed to pesticides. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this review.

a) *The analysis discussed in Appendix A of the white paper used empirical toxicity data for AChE inhibitors to evaluate the predictions of specific QSAR tools. Please discuss the applicability of the conclusions based on AChE inhibitors to other pesticide MOAs. Although not part of the current analysis, please comment on the ability of SAR tools to quantitatively predict the toxicity of chemicals with non-specific MOAs for sensitive animals (e.g., pesticide degradates and some herbicides with a narcosis MOA), that may be within the domain of applicability of these tools.*

b) *Please comment on potential modifications to the existing versions of ECOSAR, TEST and the OECD toolbox that could improve their quantitative capability relative to estimating acute toxicity values for aquatic animals exposed to specific pesticides. For example, expansion of QSAR models within existing tools to better include pesticide active ingredients within their domain of applicability, expansion of the model training sets to cover taxa associated with data gaps, development of models based on MOA, inclusion of parameters other than the logarithm of the octanol-water partition coefficient (i.e., log P) to describe toxicity.*

Panel Response Summary

The EPA presented their findings of the performance of three commonly used predictive methods based on (Q)SARs, chemical groupings and read-across approaches in Appendix A of the white paper. The investigation was a thorough analysis of how well the three techniques, Ecological Structure Activity Relationships (ECOSAR), Toxicity Estimation Software Tool (TEST), and the Organization for Economic Cooperation and Development (OECD) QSAR Toolbox, predicted toxicity for pesticide and pesticide-like compounds for which data were obtained, but were not necessarily already known to the models. The first two methods considered are based on (Q)SARs developed to predict the toxicity of industrial chemicals regulated by EPA under the authority of the Toxic Substances Control Act (TSCA). The OECD QSAR Toolbox has been developed as part of an international project to allow for grouping of compounds to facilitate read-across; the tool was primarily developed for the type of compounds that would be considered under TSCA. The Panel appreciated the EPA's methodical and careful analysis of the investigation, as well as the high quality and clarity of the documentation and the presentations during the meeting. The results indicated that the current (Q)SAR and grouping methods investigated could ultimately be appropriate to predict the toxicity of pesticides, but were lacking because the pesticides being considered and their mechanisms of action were outside of the applicability domain of the current models. The Panel agreed with the findings of the analysis. In other words, the models were not developed with data sets that are representative

of pesticides and the acetylcholinesterase (AChE) inhibition mechanism in particular, and they were not built using descriptors appropriate for modeling the toxicity of these compounds. Overall, the Panel agreed with the conclusions presented by the EPA that more work is required in the development of QSAR models, specifically in the sourcing of high quality toxicity data sets and the use of molecular descriptors that are more appropriate to the chemical classes and mechanisms of action considered. The Panel also recommended use of adverse outcome pathway (AOP)-based models to address chemical interactions and mixtures. The Panel noted the strong possibilities for the use of tools developed for compounds with non-specific toxicity, e.g., narcosis used to predict the toxicity or effects of degradates. The key for successful use of all these models is in the definition and appropriate use of their applicability domain. Finally, the Panel strongly encouraged the EPA to publish the findings of the evaluation of (Q)SARs to predict the toxicity of pesticides.

Charge Question 2: Web-ICE for Predicting Acute Toxicity for Aquatic Animals

a) *Appendix B of the white paper includes a review of the utility of Web-ICE for quantifying sensitive acute toxicity values for aquatic animals exposed to pesticides. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this review.*

Panel Response Summary

The Panel considered the conclusions made by EPA regarding the utility of Web-ICE in Appendix B to be reasonable. The Panel agreed that there were numerous strengths to this program especially in the potential ability to predict toxicity of a chemical to a species within the same taxonomic group and of similar life history for which large data gaps exist, for example for a threatened or endangered species. The fact that the program is transparent and freely available on the internet increases its utility. However, the program is limited by only incorporating acute toxicity responses and by not being applicable to predict toxicities across different phyla. In order to be useful to predict chronic toxicity, the program will have to incorporate mode of action of the chemicals, specifically geared to defining AOPs that are linked to endpoints relevant to growth and reproduction. Additional recommendations to increase the utility of the program included adding estimates of confidence to the predictions, eliminating data redundancies that currently reduce functionality of the program, and potentially incorporating (Q)SAR approaches for specific chemicals and comparing species with similar life history strategies. Specific strengths and weaknesses of Web-ICE were enumerated.

b) *Please comment on potential modifications to the existing version of Web ICE that could improve its ability to predict the acute toxicity values of pesticides to different species of aquatic animals. For example, expansion of the tool to consider MOA, addition of data for infrequently tested species that may fulfill the data requirements defined in the 1985 guidelines.*

The Panel had a number of suggestions to improve the existing version of Web-ICE, version 3.1. Most of the suggestions applied to improving the prediction of acute toxicity, although some

improvements were suggested that included predictions of chronic toxicity among species. The main suggestions were:

- Systematic evaluation of each of the minimum data requirements (MDRs) to determine if every species listed is needed for every chemical, mode of action (MOA), and system;
- Inclusion of chronic effects, particularly those related to growth and reproduction and consideration of life history;
- More directed data collection that fills in points along the MOA and adverse outcome pathways, particularly for those chemicals that have specific effects; and,
- Development of Web-ICE for algal and plant species.

Charge Question 3: Use of SSDs to Estimate HC5 using Varying Amounts of Data

a) Appendix C of the white paper includes a description of the analyses of SSDs. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this analysis.

b) Please comment upon additional analyses relative to SSDs that could be considered by USEPA that may result in improved confidence in estimates of HC5 values using various amounts of data that are typically available for registered pesticides.

Panel Response Summary

The Panel concluded that the Agency's analysis described in Appendix C of the white paper is a logical and well thought out consideration of the adequacy of available data, MDRs, and methods for estimation of the HC5 using species sensitivity distributions (SSDs). Both the resampling and distributional analyses show that imposing taxonomic requirements on the sampling can have a substantial impact on the HC5 estimate and can lead to conservative estimates for some combinations of small sample size and estimation method. The constrained sampling represented by the MDRs or pesticide data requirements results in 95% confidence intervals that are almost guaranteed to contain the true HC5. In general, the methods used are based on empirical relationships derived from statistical analyses and not on a mechanistic or biological model. Without greater mechanistic or biological understanding, it is not possible to explain why specific combinations of assumed distributional form, estimation method, sample size and sample constraint are unable to produce acceptable estimates of the HC5. As sample size decreases, there are insufficient data available for estimation of the HC5.

The Panel commented that the maximum likelihood, method of moment and graphical estimation approaches evaluated in Appendix C have both positive and negative features. The Panel discussed and elaborated on the study results and provided additional insights into why certain methods might work better in certain situations:

- The method used to estimate the parameters of the species sensitivity distribution (SSD) depends on the distributional form assumed, with some distributions allowing easy

parameter estimation via a log10 transformation and least square methods, and other distributions allowing only method of moment estimates or requiring programmed maximum likelihood algorithms. Because of this, it is difficult to determine whether underestimation of HC5 is due to assuming the wrong distributional form for a set of data or due to failure of the available parameter estimation methods.

- The Panel reiterated the need for high quality and validated quantitative data to support HC5 estimation with strong discouragement of study designs that are only capable of providing NOEC or LOEC values.
- Better performance of the triangular distribution is likely the result of data collection under minimum data requirements that results in a more uniform than random sampling of the distribution of acute outcome values.
- The model averaging section was difficult to understand. Small sample sizes result in Akaike information criteria (AIC) values that do not differ among fitted parametric distributions resulting in equal weights in the model averaging. The Panel suggested that weights based on goodness-of-fit related to the graphic estimation approach might provide better results.

The results presented in Appendix C demonstrated that the use of SSDs to estimate the HC5 is improved when MDRs are met. The Panel made the following comments and suggestions for future study directions including, but not limited to:

- The EPA should move away from chemical-by-chemical approaches to approaches centered on classes of chemicals sharing common toxicity characteristics such as the use of similar MOAs and AOPs.
- Use of order statistics approaches may not be a productive source of alternative estimation methodologies.
- While Bayesian modeling and estimation approaches (e.g., empirical Bayes methodologies) provide a better framework for considering MDRs and taxonomic representativeness, interpreting the results poses challenges that need to be addressed. Both data and chemical characteristics may drive the success of a particular SSD distributional model and estimation method. EPA should examine more chemicals, more MOAs and smaller, less complete datasets before the full utility of this kind of analysis can be decided.
- EPA has not formally and systematically assessed the adequacy of using subsets of MDRs or of supplementing MDRs with species which require special consideration (e.g., endangered species, most sensitive species). The Panel recommended this be done sooner than later and recommended certain species be studied first.
- Exposure and toxicity modifying factors, such as salinity, pH, hardness, etc., should be included in studies that measure acute effects to enable data collected under controlled laboratory conditions to be translated to the real world.
- EPA should begin to consider how chronic endpoints will be included in SSD analyses.

Charge Question 4: **Extrapolation Factors for Estimating Acute HC5s with a Specified Level of Confidence**

a) Please comment on the extent to which consideration of MOA decreases uncertainty associated with HC5 values that are estimated using EFs.

Panel Response Summary

EPA tested the validity of the MOA approach using AChE inhibitors and narcotic chemicals (Appendix D). The Panel concluded that these examples support using MOA to decrease uncertainty associated with HC5 values, especially those estimated using EFs. The Panel recognized that invertebrates (e.g., arthropods) were more sensitive than fish to the acute effects of AChE inhibitors. This was the case for all six AChE inhibitors tested. The Panel encouraged EPA to further develop the MOA approach for chemicals with other MOAs, e.g., synthetic pyrethroid insecticides. Evaluation of sets of chemicals with different MOAs will provide a more robust dataset in which to evaluate uncertainty in estimated HC5s. The Panel cautioned that close attention should be paid to the most sensitive species which may differ for different chemicals with the same MOA. The Panel also stressed the importance of considering sensitivity differences in resident versus non-resident species, intra- and inter-species differences, and protection of endangered or threatened species as the MOA approach is developed. The Panel recommended an explicit evaluation of the importance of each species to meet a MDR. That is, inclusion of some species may make little difference in estimating the HC5. In many cases, the Panel indicated that fewer species might be needed to meet a MDR.

b) Please comment on the strengths and limitations of both the resampling and distributional methodologies for developing MOA-specific EFs, and provide suggestions on their further development and application to other MOAs. Please provide suggestions for deriving EFs for chemicals with MOAs that have limited amounts of empirical toxicity data.

Panel Response Summary

The Panel evaluated both the resampling (Host et al. 1995) and distributional methodologies for developing MOA-specific EFs and concluded that the distributional approach may have some advantages over the resampling approach. The Panel suggested that the EPA maintain the flexibility of using both approaches to address the variety of MDR data sets for different chemicals and different MOAs.

The Host et al. (1995) resampling method provides a process using random estimates of both the EF and the HC5, but these two estimates have conditional probabilities that may contribute to great uncertainty to estimates made using this approach. The new distributional approach, which modifies the Host et al. (1995) method, only estimates the EF by random resampling, uses the entire data set within the MOA to estimate the HC5, and is not confounded as much by this conditional probability and compounded error issue of the original Host et al. (1995) method. Thus, the distributional method should result in less uncertainty than the resampling method and also has the advantage of being probabilistic.

The Panel concurred with EPA's findings specific to the AChE inhibition MOA. Panelists unanimously agreed that more chemicals representing more MOAs and AOPs should be analyzed.

The Panel concluded that the distributional approach was more flexible than the resampling approach and hence might be considered the preferred approach for exploring MOAs other than AChE inhibition. However, the decision on which approach works best for a particular situation depends on the characteristics of available data and best professional judgment.

Several panelists agreed that the EPA should continue to examine other conditions for derivation of EFs. For example, the EPA could examine the effect of excluding species whose data suggest that they are much more sensitive than the next most sensitive species. Such extreme cases tend to have high influence on the estimated EF. EFs might also be based on the median rather than on a tail percentile in order to minimize the effect of influential single data points. The EPA may also consider combining data from similar chemicals regardless of their MOA to get some idea of the range of possible EFs when assessing a chemical having little available toxicity data or unknown MOA. A narrow range of EFs would indicate that exact knowledge of the MOA may not be that important. Finally, the Panel reiterated the importance of estimating and reporting confidence limits for EFs.

c) *Please comment on the strengths and limitations of the different approaches in the Host et al. (1995) and the current work for assigning reference HC5s and summarizing EF distributions.*

Panel Response Summary

The Panel considered the strengths and limitations of the different approaches in the Host et al. (1995) method and the distributional method for assigning reference HC5s and summarizing EF distributions. Overall, the Panel found the distributional method supports the approach of Host et al (1995) in that it is possible to use data sets that are lacking in terms of satisfying aquatic life criteria MDRs to derive reasonable estimates of the HC5 including the uncertainty of these estimates. The reference HC5 value used in Host et al. (1995) was the final acute value (FAV) calculated for each of the N=8 species data sets that satisfied the aquatic life criteria MDRs as opposed to the actual FAV for the complete data set for a chemical in calculating EFs. The EFs derived by Host et al. (1995) have a distribution that is shifted up (larger value) and is more spread (higher variance) than that computed from the distributional method. In the distributional method, the HC5 value of complete data sets was used to calculate EFs which is more appropriate and has lower variance than the Host et al. (1995) method.

The smaller mean EF value derived from the distributional method results from the fact that it is sampled from a larger set of data – one that is more likely to include genera with intermediate sensitivities and uses a reference HC5 which is constant in these analyses. EPA suggested that the EFs calculated using this approach were "provisional," but informative in defining methodologies and procedures for selecting EFs. The larger variance in the results from the Host et al. (1995) approach defines the reference EF in terms of the minimum of the N=8 dataset. The HC5 of the individual N=8 datasets is a random variable that will vary dramatically from sample

to sample adding to the overall variability in the EF values by as much as a factor of two for AChE-inhibiting chemicals. Other factors considered by the Panel were the importance of reporting confidence limits for EFs, adequate data for salt water species, and the effects of climate change and physicochemical water quality (pH, eutrophication, temperature and salinity) as important modifiers of ecotoxicity.

Charge Question 5: Use of Acute to Chronic Ratios (ACRs) to Estimate Chronic Toxicity Endpoints or HC5 Values for Acutely Sensitive Species

a) *Please comment on the strengths and weaknesses of using chemical-specific ACRs to estimate chronic effect thresholds for other species and taxa.*

Panel Response Summary

The ACR approach illustrated the ability to assess the importance of factors ranging from fish vs. invertebrate species sensitivity differences, freshwater vs. saltwater species sensitivity differences and differences for organophosphates vs. carbamates in developing ACRs. The major strength of employing the chemical-specific ACRs approach is that it likely represents the most robust concept at the present time to address chronic toxicity data scarcity issues (Kenaga 1982, Sloof et al. 1986, Rand 1995, Länge et al. 1998). However, it may be warranted to simply perform additional chronic studies instead of estimating chronic thresholds particularly for mechanisms/MOAs, organisms and chronic responses that are data poor at the present time. Uncertainties for such limited datasets may be too large to overcome without additional chronic adverse outcome data. The Panel indicated that the practice of generating a mean ACR value across all available chronic endpoints, which may result from different AOPs within and among species, is not appropriate. For example, fish early life stage and life cycle values were combined for this analysis, albeit to illustrate the relative greater sensitivity of invertebrate acute response thresholds than fish chronic thresholds for AChE inhibitors.

The AChE inhibitors used in this exercise represented an excellent initial effort due to the amount of data available for this common MOA. Future studies should employ AOPs to identify appropriate chronic endpoints in various organisms prior to ACR derivation, particularly for specifically acting chemicals. For example, there is a need to assess more AOPs associated with pyrethroid insecticides (Ankley et al. 2010).

b) *Please comment on the strengths and weakness of applying "default" ACRs derived from other chemicals to extrapolate from an acute HC5 to a chronic HC5, including the relative merits of values derived for MOA-specific (e.g., for AChE inhibitors) or more generalized (e.g., Host et al. 1995) distributions.*

Panel Response Summary

Application of "default" ACR values is necessary to account for uncertainties based on the paucity of available chronic toxicity data for most industrial chemicals. When possible, however, MOA-specific ACR values, which are influenced by various chronic AOPs, should be developed by endpoint in major taxa (e.g., macrophytes, algae, invertebrates, vertebrates) using AOPs to

avoid under- or over- estimation of chronic thresholds. The Panel emphasized the point that employing default ACR values and those derived by averaging ACR values across chronic endpoints within and among species, is not recommended, particularly for biologically active molecules because various chronic responses likely result from different AOPs. At best, in the absence of these considerations, the Panel recommended the development and application of uncertainty factors to ACR values calculated across endpoints and species for biologically active chemicals. ACR values should only be used for species that employ similar life history strategies (for example short life span, fast growth, high reproductive effort vs. long life span, slow growth and reproductive effort spread over several years). Taxonomic differences during AOP development to account for differences in sensitivity of freshwater and saltwater fish to a chemical and invertebrate sensitivity to the same chemical should be considered.

The Panel was supportive of the generic approach proposed in the presentation made by the public commenter representing CropLife America to develop a hierarchy of preferred data attributes during ACR development.

c) *Are there other methods for estimating a chronic HC5 that the panel believes would be technically superior to ACR-based approaches? For example, TCE models (described in USEPA 2010a), distributional approaches (e.g., de Zwart 2002, Douboudin et al. 2004; described in USEPA 2010b).*

Panel Response Summary

The Panel commented on the use of distributional approaches, TCE models, and other approaches for estimating a chronic HC5 other than an ACR-based approach. The Panel indicated that distributional approaches showed great promise, but data availability will inherently limit the broad implementation of this approach. Further, due to potential differences in chronic AOPs among species, ACR values can vary greatly among species and chronic endpoints selected. TCE models may be useful for some chemicals, but the robustness of the models will be limited when chronic MOAs differ from acute MOAs. Another proposed approach was to explore the utility of deriving ACR values from differences between centiles from species and chronic endpoint specific toxicity distribution for chemicals with a common MOA as demonstrated for AChE inhibitors (Berninger 2011, Williams et al 2011). But here again, data availability limits broad application of these techniques at this time. Use of "omics" and "*in vitro*" strategies such as being developed by the National Toxicology Program would be useful to focus testing strategies.

Charge Question 6: Estimating the HC5 for Aquatic Plants

a) *Appendix F of the white paper includes an analysis of the use of EFs and SSDs combined with the FIFRA-5 to estimate the HC5 for aquatic plants. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this analysis.*

b) *Please comment on other aquatic plant species or test endpoints that may be considered in order to generate more certain estimates of the HC5 for aquatic plants.*

c) *Please comment on the strengths and limitations of normalizing plant toxicity endpoints to a standard metric (i.e., EC50 for growth rate).*

d) *Please comment on the extent to which consideration of MOA may decrease uncertainty associated with HC5 values for aquatic plants.*

Panel Response Summary

The Panel appreciated the tremendous amount of work that went into these analyses. In general, the Panel believes that this work has laid the foundation for aggressively moving towards achieving the goal of harmonizing efforts, requirements and conclusions within the EPA. In response to this charge question, the Panel provided the following recommendations to further advance the Agency's efforts with respect to the protection of aquatic plants.

- Develop MDRs for freshwater and saltwater plants.
- Verify the uncertainty (i.e., over- or under-protection) inherent in the use of certain extrapolation factors for a given number of species in a SSD.
- Develop and implement a strategy for systematically replacing data that were collected using suspect methods or analytical procedures, using nominal values or not based on curve-fitting endpoint estimates (e.g., NOECs). However, the Panel recognized that in cases in which historical toxicity test data may provide the only relevant toxicity test results for specific chemicals of concern, then the EPA should use their best professional judgment regarding the inclusion of such data in the risk analysis.
- Characterize the influence of nutrients on plant toxicity test results both in the standard laboratory bioassays used to generate data for regulatory purposes (issues of non-uniformity among bioassays and investigators) and in the field (issues with nutrient enrichment in many freshwater and saltwater systems).
- Consider the addition of aquatic macrophyte toxicity tests with traditional (e.g., shoot growth) and non-traditional but potentially more sensitive (e.g., root growth) endpoints. Addition of protocols and endpoints should account for differences in the MOA and mechanism of action of specific herbicides and fungicides.
- Consider the addition of a marine macrophyte toxicity test protocol, and potentially other species (e.g., to develop a FIFRA-5 corollary for marine organisms) in the MDR development.
- Determine ecologically relevant endpoints (i.e., those that determine population dynamics, such as growth and reproduction in animals) for aquatic plants and couple them with animal (herbivore) endpoints to understand potential secondary effects (e.g. loss of primary producers resulting in reduced food for herbivores).
- Initiate a systematic research effort to determine the appropriate effects concentration (HCx) to protect plant community structure and function.
- Incorporate knowledge of AOPs in test species selection and data interpretation to reduce the large uncertainties that exist in this process.

Additional Comments

The Panel provided additional comments following the discussion of all of the Agency's charge questions. The Panel commended the Agency scientists for their excellent work and strongly recommended that their research be presented at national and international scientific meetings and published in peer reviewed journals. Threshold response values that are being developed based on limited databases have a relatively high degree of uncertainty. There should be an effort to field validate these values to see whether they are protective of populations and communities. The Panel strongly supported the development and application of AOP approaches for use in ecological risk assessment. Similarly, sublethal responses (e.g., biomarkers) should distinguish between measures of exposure and of effects. These are appropriate for use in the early stages of the ecological risk assessment process, problem formulation and screening level assessment, but only sublethal measures indicating adverse outcomes should be used in the final detailed assessment. Finally, the Panel recommended periodic scientific review and updating of the Agency's documents, guidelines and benchmarks as stated in the 1985 Guidelines' "Good Science Clause."

DETAILED RESPONSES TO CHARGE QUESTIONS

Charge Question 1: (*Q)SAR Tools to Predict Acute Toxicity Values*

Effect characterizations in both United States Environmental Protection Agency's (USEPA) Office of Water (OW) and Office of Pesticide Programs (OPP) rely on toxicity estimates for various taxa. The USEPA investigated the use of some readily available Quantitative Structure Activity Relationship (QSAR) tools ***(****described in* ***Appendix A)*** *to predict acute toxicity values to represent sensitive taxa which could potentially fill gaps in available empirical toxicity data. Three publicly available QSAR models were examined, including: Ecological Structure Activity Relationships (ECOSAR), Toxicity Estimation Software Tool (TEST), and the Organization for Economic Cooperation and Development (OECD) QSAR Toolbox. These tools are increasingly used by USEPA risk assessors to determine the potential toxicity of various chemicals, including pesticide degradates. The white paper's analyses of ECOSAR, TEST and the OECD Toolbox suggest that, at this time, 1) the chemical domains for most models within these QSAR tools do not include sufficient data for pesticide active ingredients and therefore these models are not currently robust estimators of the toxicity for modes of action (MOAs) that are unique to pesticides (e.g., acetylcholinesterase (AChE) inhibitors), 2) current models are generally populated with data for species for which toxicity data are typically submitted under FIFRA, and generally, could not be used to fulfill other taxonomic data requirements defined in the 1985 guidelines, and 3) more toxicologically relevant molecular descriptors and structural alerts need to be identified for use in building models for MOAs associated with pesticide active ingredients.*

Panel Response

a) *Appendix A of the white paper includes a review of the utility of ECOSAR, TEST and the OECD toolbox for quantifying sensitive acute toxicity values for aquatic animals exposed to pesticides. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this review.*

The EPA presented their investigation into the performance of three commonly used predictive methods, (Q)SARS, chemical groupings and read-across approaches, in Appendix A of the white paper and during the oral presentations. The investigation was a thorough analysis of how well the three techniques, ECOSAR, TEST and the OECD QSAR Toolbox, predicted toxicity for compounds for which limited data were obtained but were not necessarily tested, using the computational methods. The SAP appreciated the methodical and careful analysis of the investigation, as well as the high quality and clarity of the documentation and presentation. The SAP was in agreement with the main findings of EPA's analyses. The results indicated that the current (Q)SAR methods investigated could ultimately be appropriate to predict the toxicity of pesticides, but were lacking at the current time due to the pesticides considered being outside of the applicability domain of the models considered. In other words, the models are not developed on data sets representative of pesticides, and the AChE inhibition mechanism in particular, and they are not built using descriptors appropriate for modeling the toxicity of these compounds. The SAP agreed with the conclusions presented by the EPA that more work is required in the development of models, specifically in the sourcing of high quality toxicity data sets and the use of molecular descriptors that are more appropriate to the chemical classes and mechanisms of

action considered. The SAP strongly encouraged the EPA to publish the findings of the evaluation of (Q)SARs to predict the toxicity of pesticides.

Specific Comments

The Panel provided the following specific comments on the analysis performed and reported in Appendix A.

- The EPA captured and organized the existing toxicity data very well. This provided confidence in the findings and results of the analysis, as well as providing a basis for (Q)SAR development.
- For estimates from ECOSAR, TEST, etc., the Panel suggested that an annotated (i.e., with significant outliers) plot of predicted versus observed toxicity be added to Appendix A just as was used in the Agency's presentations during the meeting.
- The concept of a controlled vocabulary for molecular descriptors, as presented by the EPA, is strongly supported by the Panel. For models to be useful, "standard" freely available molecular descriptors that are well defined and unambiguous are needed.
- Well-defined applicability domains are needed for successful use of all (Q)SARs and predictive models. These can be defined in terms of chemical structure, physico-chemical, mechanistic and metabolic space (Dimitrov et al 2005).
- The use of "docking studies" for receptors or target proteins is recommended for modeling specific MOAs where receptor interactions occur (Biesiada et al. 2011; Garcia et al. 2010; Wu et al. 2010; Celander et al. 2011). This could assist in the search for commonalities across species and help to model binding constants and potency of the different contaminants.
- In order to assist (Q)SAR modeling, the fairly standard practice of plotting toxicity against log P could be informative for pesticides. In particular (where available) the baseline narcosis toxicity (Q)SAR could be drawn on the plot for the individual species. This would enable the determination of whether the pesticide has significant excess toxicity and if there is any significant trend of toxicity with log P that would be of use in the ECOSAR and category formation.

The Panel identified the following strengths and weaknesses in each of the three (Q)SAR methods: ECOSAR, TEST, and the OECD QSAR Toolbox.

1) ECOSAR

The ECOSAR software is freely available as a download from the EPA web-site. It provides predictions of acute toxicity for a large number of chemical classes.

The following strengths of ECOSAR were identified:

- ECOSAR is a well-established (Q)SAR program and approach for predicting toxicity. It has undergone many years of development, and is continuing to undergo development, within the EPA. Various analyses (de Haas et al 2011; Reuschenbach et al 2008) have shown it to be robust and predictive of the toxicity of industrial chemicals regulated by

EPA under the Toxic Substances Control Act (TSCA). It is commonly used outside of the US (e.g., in the European Union for Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) submissions as a computational tool to predict toxicity. Each version of ECOSAR had improved predictivity versus the previous one. For example, version 1.1 was shown to define chemical groups better, with regard to pesticides than version 1.0. This indicates that relatively minor alterations and adaptations to the (Q)SARs can bring about significant improvement in ECOSAR.

- The ECOSAR software is freely available as a download. This means it can be used by any stakeholder to obtain predictions. The models are transparent, i.e., there are unambiguous algorithms associated with the original data sets (with the exception of confidential business information) in compliance with the OECD Principles for the Validation of (Q)SARs (OECD 2007). As such, the models are amenable to the easy creation of documentation to support a prediction, e.g. through the (Q)SAR Model Reporting Format (QMRF).
- The ECOSAR software is based extensively on models derived from logarithm of the octanol-water partition coefficient (log P) values. This assists with the transparency of the models. The log P values are derived from a significant database of high quality measured values or a robust estimation method, KOWWIN, freely available software in EPISUITE provided by the EPA, used to predict log P. Log P is known to have strong correlations with the acute aquatic toxicity of numerous compounds, particularly within groups of compounds and/or mechanisms of action.
- ECOSAR essentially uses a grouping approach to form groups or categories of chemicals. Grouping of compounds is appropriate to develop models (read-across / SAR / (Q)SAR) for specifically acting pesticides.

The following weaknesses of the ECOSAR software were identified:

- The Panel agreed with the comments from the EPA that ECOSAR version 1.0 was not developed to predict the toxicity of pesticides. Specifically, the definition of groupings is not sufficient for the prediction of pesticides and some, relatively general groups, e.g. organophosphates, should be better defined. Whilst ECOSAR version 1.0 performed poorly in the prediction of toxicity, some improvement was seen in version 1.1. The improvement was specifically in the area of definitions of groups of compounds relevant to pesticides and the AChE MOA. Despite this, the improvements need to be supported by further toxicity data for pesticides.
- Associated with the problem of grouping is that of “selecting” appropriate classes of compounds for the predictions. The Panel agreed that EPA used appropriate “expert opinion” in the expert analysis reported in Appendix A.
- The Panel pointed out that it is known that the (Q)SARs in ECOSAR have been developed from very few toxicity data (Kaiser et al 1999). This may lead to problems with the statistical analysis and reduce the predictive capability within that group. This is a pertinent issue given the paucity of data in ECOSAR for pesticides.
- While there may be drawbacks in using a small number of data to develop a (Q)SAR, the approach taken by the EPA is pragmatic. The (Q)SARs are well documented so the user will be aware of the statistical issues related to a particular model and can use this information to determine if a particular prediction is appropriate.

- The input into the ECOSAR software is via Simplified Molecular Input Line Entry Specification (**S**MILES strings). These are 2-D in nature. The Panel indicated that there may be a move to more sophisticated methods of molecular input, e.g., IUPAC International Chemical Identifier (InChI) keys / codes. Building and applying (Q)SARs in this way would assist in making better predictions for chiral molecules and stereoisomers.
- With regard to the performance of the ECOSAR software as reported in Appendix A, the predictions for some compounds often indicated them to be less toxic than indicated by the definitive (experimental) studies. This suggests that ECOSAR in its present form is under protective of sensitive species.
- Some toxicity data included in the analysis of predictivity were for compounds that may have been included in the (Q)SARs within ECOSAR. Ideally, these should be removed from the analysis, or separated from the true validation data, to give a representative view of the predictivity of ECOSAR.

2) TEST

The TEST software is freely available as a download from the EPA web-site (http://www.epa.gov/nrmrl/std/qsar/qsar.html#TEST). It provides predictions of acute toxicity (as well as other endpoints) for a large number of chemical classes. The Panel listed a number of strengths and weaknesses of TEST.

The following strengths of TEST were identified:

- The TEST software utilizes more molecular descriptors than just log P. This may mean that it is better able to predict the toxicity of pesticides.
- The TEST software performs a consensus approach from a variety of models. Such approaches are known to enable more accurate predictions of toxicity (Matthews et al. 2008).

The following weaknesses of TEST were identified:

- A variety of (Q)SAR methodologies and approaches used in the TEST software go beyond regression analysis. This may result in a lack of transparency (in the models) as unambiguous algorithms are not immediately accessible as well as reducing confidence in the predictions as their basis may not be clear. In addition, these methods are not necessarily proven for predicting aquatic toxicity of pesticides at the current time.
- Although, the TEST (Q)SAR models appear to be based on large databases of toxicity values, very few, if any of the datasets include pesticides. This means that the model will be biased towards the types of compounds within the dataset. Most data sets of toxicity values are dominated by narcotics; therefore, the model(s) may be driven by those most common chemical classes. As such, the ability of the (Q)SAR to make predictions for pesticides will be reduced.
- The multivariate statistical approaches employed by TEST (i.e., many descriptors) are less likely than ECOSAR or the OECD QSAR Toolbox to be able to develop statistically valid (Q)SARs for small data sets such as may be available for pesticide molecules.

- TEST methods indicated more molecular descriptors were used, but it was unclear how the "unified file" for MOA was constructed. ASTER was not accessible via the website provided. IRAC had relevant MOA information, but it was unclear how this was used with the somewhat older and more general descriptions of Russom et al. (1997). Further the method does not include (is not based on) MOA or chemical categories. The "unified file" was the term used in the appendix and was supposedly the file which incorporated MOA into the TEST model. ASTER was the acronym used in the text for a web-based program that incorporated MOA into the TEST system. Attempts were made to evaluate the MOA list at the website provided in the documentation, but access was denied. IRAC was another website that showed MOA that were utilized for the model (it was the only one that worked). It showed a list of 8-10 MOA which were much more specific than the 4-5 which were used in Russom et al. 1997 which were primarily 1-3 subtypes of narcosis.

3) OECD QSAR Toolbox

The OECD QSAR Toolbox (the Toolbox) is a freely available software application available as a download from the OECD web-site (http://www.oecd.org/document/54/0,3746,en_2649_34379_42923638_1_1_1_1,00.html and http://www.qsartoolbox.org/). It is primarily a tool for chemical grouping with the ability to provide predictions of toxicity from read-across, trend analysis and (Q)SAR analysis.

The following strengths of the OECD QSAR Toolbox were identified:

- The OECD QSAR Toolbox (the Toolbox) is freely available and under development, until the end of 2012 at least. It is complete with a variety of methods, and profilers, to group chemicals together and perform read across and limited trend analysis / (Q)SAR analysis.
- The Toolbox will become increasingly important for grouping of compounds based on mechanisms (and modes) of action. Therefore it could have considerable use for the prediction of the toxicity of pesticides.
- The Toolbox has access to large databases of toxicity values. However it appears that none at the moment are specific for pesticides.
- As demonstrated in the analysis reported by the EPA, the Toolbox has a (limited) ability to calculate molecular descriptors on the fly and therefore develop appropriate (Q)SAR models for pesticides.

The following weaknesses of the OECD QSAR Toolbox were identified:

- As noted by the EPA, the Toolbox has not been developed with pesticides in mind. It is very much a tool to make groupings for industrial chemicals (i.e., those covered by TSCA). Despite this potential drawback, this does not preclude consideration of pesticides by the Toolbox should appropriate profilers and data be made available.
- It is unclear at the current time how well the Toolbox defines the mechanism of action of the AChE inhibitor class. In other words, there are a variety of groupings relevant to the

organophosphates (e.g. as structural fragments) but they may require further definition and clarification – particularly to be relevant at the mode and mechanism level. It is noted that it is possible to develop bespoke "profilers" in the Toolbox (and in other technologies). These could be developed for specific pesticide mechanisms of action. Should this be done, they could be donated to the Toolbox through US representation at OECD. The development of pesticide specific mechanism profilers may have the added advantage of encouraging use of this tool for the evaluation of pesticides.

- As an example of the development of appropriate groups, peripheral sites of oxidation (i.e., thioether linkages) could also be flagged as structural identifiers for both groups of AChE inhibitors. This may explain the "outlier" (i.e., aldicarb) reported by the EPA.
- As acknowledged in the documentation, the (Q)SAR models developed by the EPA from the Toolbox are relatively poor – both in terms of statistical fit and predictivity. A possible reason for poor performance is (despite the name) the Toolbox is not a sophisticated method to develop (Q)SARs. If (Q)SAR development is intended in future analyses, it would be better to form the group of chemicals and identify toxicity data within the Toolbox, extract the structures and data, and use other tools to create models. There are a variety of increasingly sophisticated software tools for physico-chemical / molecular descriptor calculation and statistical analysis. Some tools are freely available, e.g., Chemistry Development Kit (CDK available from http://sourceforge.net/projects/cdk/), Molecular Orbital Package (MOPAC available from http://openmopac.net/index.html) etc for descriptor calculation, the R Project for Statistical Computing (R available from http://www.r-project.org/), Waikato Environment for Knowledge Analysis (Weka available from http://www.cs.waikato.ac.nz/~ml/weka/), etc., or may include commercial software e.g. OASIS (available from http://oasis-lmc.org/?section=software&swid=12), Comprehensive Descriptors for Structural and Statistical Analysis (Codessa available from http://www.semichem.com/codessa/default.php), Molecular Operating Environment (MOE available from http://www.chemcomp.com/software.htm), Comparative Molecular Field Analysis (CoMFA available from www.tripos.com), etc. (Please note this is not a comprehensive list of software and other tools and products are available.)
- All (Q)SAR models use molecular descriptors derived from two-dimensional structure (although it is appreciated that the molecular orbital properties are calculated from an optimized three-dimensional structure). There would be an advantage in the use of three dimensional descriptors that take chirality into account given the stereoselectivity associated with various molecular targets (i.e., AChE). In addition, bioactivation rate constants for organophosphate pesticides using Km values from several species should provide some reduction of uncertainty. If we know how much of the OP is being enzymatically converted to the oxon, then relative comparisons to toxicity may be stronger. As noted by the EPA, it was difficult to perform appropriate toxicity data selections in the Toolbox, particularly with regard to filtering and quality. The Panel agreed with the EPA that a taxonomic hierarchy and mechanism for unit standardization is needed.

b) *The analysis discussed in Appendix A of the white paper used empirical toxicity data for AChE inhibitors to evaluate the predictions of specific QSAR tools. Please discuss the applicability of the conclusions based on AChE inhibitors to other pesticide MOAs.*

Although not part of the current analysis, please comment on the ability of SAR tools to quantitatively predict the toxicity of chemicals with non-specific MOAs for sensitive animals (e.g., pesticide degradates and some herbicide with a narcosis MOA), that may be within the domain of applicability of these tools.

The Panel agreed that the conclusions reached by EPA in Appendix A serve as guidance for the development and application of (Q)SARs for specifically acting toxicants, e.g., pesticides. These findings and approaches to predict toxicity are relevant to other specifically acting pesticides; thus, they could be applied to other mechanisms of action. The extension to other mechanisms of action will require modeling of the relevant receptors, for instance binding pockets and crystal structures when available which will help focus the (Q)SAR tools. The modeling of specific mechanisms can be supported using information from qualitative screens such as with Toxicology in the 21st Century (Tox21), a collaborative effort among several Federal Agencies to develop innovative chemical testing methods that characterize toxicity pathways (http://epa.gov/ncct/Tox21/). This could provide information required in subsequent evaluations (i.e., sensitive taxa and endpoints), for example, if a compound does not indicate acute toxicity, but binds to the estrogen receptor (ER), then assessment studies should focus on reproduction or life cycle assessment of vertebrates rather than invertebrates.

In order to improve (Q)SAR predictions for specific mechanisms based on receptor binding, more specific parameterization of binding metrics (e.g., molecular descriptors associated with AChE inhibition) is needed. This will be the same for other mechanisms, i.e., better descriptors relating specifically to the mechanism of action. The (Q)SAR models reported in Appendix A, as they stand currently, are not likely useful for quantitative predictions of toxicity from AChE inhibition. In addition, further information may be needed, e.g., some assessment of the activation of an organophosphate to the oxon, or in some cases carbamates, (e.g., aldicarb). In general terms, (Q)SAR approaches may work best for well-characterized receptors (i.e., AChE, gamma-aminobutyric acid (GABA), ecdysone, estrogen receptor (ER) and aryl hydrocarbon receptors (AhR)) and may predict binding and subsequent activation. Further, it will be imperative to link these to apical endpoints of reproduction, growth, and survival (through an AOP framework approach) to provide a way to eventually tie effects to population dynamics.

The Panel noted that there are strong possibilities for the use of tools developed for compounds with non-specific toxicity, e.g., narcosis, to predict the toxicity or effects of degradates. The key to the success of QSAR models in this context is in the definition and appropriate use of the applicability domain. Should a compound fall within the applicability domain of a particular (Q)SAR, category or model, then it may be used to predict toxicity regardless of whether the compounds are pesticide degradates or not.

c) *Please comment on potential modifications to the existing versions of ECOSAR, TEST and the OECD toolbox that could improve their quantitative capability relative to estimating acute toxicity values for aquatic animals exposed to specific pesticides. For example, expansion of QSAR models within existing tools to better include pesticide active ingredients within their domain of applicability, expansion of the model training sets to cover taxa associated with data gaps, development of models based on MOA,*

inclusion of parameters other than the logarithm of the octanol-water partition coefficient (i.e., log P) to describe toxicity.

The Panel endorsed the conclusions presented by the EPA that the three (Q)SAR tools evaluated for the prediction of the acute aquatic toxicity could be improved to increase their performance. A key requirement, and very essential concern in the development of robust (Q)SARs, is the need for high quality and reliable toxicity data. This point was clearly made by EPA in Appendix A, i.e., the data sets of pesticide toxicity values were not as consistent as, for instance, the fathead minnow (http://www.epa.gov/ncct/dsstox/sdf_epafhm.html) and *Tetrahymena pyriformis* (Schultz 1997) databases. Current efforts to compile toxicity data for pesticides run the risk of combining together data that may show inter-laboratory variability. The Panel recommended that EPA address this issue in some way and to develop a strategy for identifying high quality toxicity data sets.

In general, there is a spectrum of (Q)SAR methods to predict toxicity from models derived from statistical analysis of large data sets that may cover many mechanisms of action (e.g., TEST) through to mechanistically derived models (e.g., those reported from the OECD QSAR Toolbox). The Panel noted that different statistical and modeling techniques will be required for the various (Q)SAR approaches. Therefore, different methods will be needed to improve the three (Q)SAR approaches presented by the EPA. Large datasets that represent many mechanisms of action are effectively modeling (bio)availability, the ability to interact (i.e., inter-mechanism effects) and the strength of the interaction (i.e., intra-mechanism potency). Molecular descriptors for all these properties should be present in a model. For single mechanisms, only two effects need to be modeled, (bio)availability and the strength of interaction. Within a group of compounds, if the strength of interaction is constant, such as narcosis interactions, then it is only (bio)availability which is important. This is why there are good correlations and (Q)SARs with log P alone. For specific and / or receptor-mediated mechanisms of toxicity, e.g., AChE inhibition, there will probably be a variation in potency of the receptor binding that should be modeled.

1) Modifications to the three (Q)SAR models

The Panel identified a number of modifications that could be made to the three (Q)SARs to improve their performance:

- Further toxicity data could be compiled to expand the domain of the (Q)SARs and thus make models more robust.
- The applicability domains of all (Q)SARs could be defined better to allow for the assessment of whether compounds for which predictions are made are within their domains.
- Models could be prepared with QMRF documentation to allow for their easier reporting and assessment.
- Data gaps in the domains of the models and compounds for which data are required could be identified or mapped. This could enable appropriate model development in areas of chemistry (or mechanisms) where information is missing.
- AOP-based models should be developed to address chemical interactions and mixtures. This approach has already been discussed in human health risk assessment

paradigms and should also be incorporated into ecological risk assessments. The Panel noted that the EPA Science Advisory Board in 2009 addressed the use of AOP for emerging contaminants including pesticides (USEPA SAB 2009). The Panel recommended that EPA update AOPs as data become available on specific chemicals.

- The Panel stated that an acute MOA may not be the same as a chronic MOA. The incorporation of MOA / AOP concepts in model development will help link chronic effects to endpoints such as growth, reproduction, survival. Such linkages are crucial for the formation and definition of chemical categories. The AOP concept may allow for targeted (non-animal such as ToxCast) testing to define the domains of a category. Further, categories (or chemical groupings) can be developed for recognized AOPs. This has the advantage that the category, with a definable applicability domain, can be linked directly to endpoints such as growth, reproduction, survival, etc.
- (Q)SARs may be developed for the same species and endpoint, but separately for different experimental conditions e.g. pH, dissolved oxygen content, temperature, salinity etc. Comparison of these QSARs e.g. by determining the relative contributions of descriptors could assist in developing more predictive models. This knowledge may be important to understand the relative effects of experimental across taxa and hence for multiple stressor type situations.
- The (Q)SARs can be compared across taxa with similar or different life history strategies, particularly when MOA is considered and there is a link to reproduction or growth.
- Modeling binding to receptor/protein targets is recommended. The results of predictions from ECOSAR and TEST show the lack of correlation with AChE inhibitors and indicate other pharmacokinetic parameters, e.g., Vmax, Km, Kd etc., may be more important for modeling.
- (Q)SARs could assist in the identification of other issues including differences between saltwater and freshwater fish species which may have different uptake mechanisms.
- Use of further descriptors in (Q)SARs should be done on a mechanistic basis, i.e., descriptors should be selected that relate to the mechanism of action.

2) Examples of descriptors used in (Q)SARs for predicting the toxicity of pesticides

The Panel provided the results of a very brief literature search for descriptors used in (Q)SARs for prediction of the toxicity of pesticides and (Q)SARs for AChE inhibition. Many different descriptors have been applied as shown in the following list. However, the list should be treated with caution as each analysis will use different descriptor sets, different training sets and different modeling approaches. Some descriptors such as the presence of atoms / hydrogen bonding group / fragments, simply represent structural diversity within a dataset and would not be expected to be important if a congeneric series was considered.

Hydrophobicity:

- Log P (log Kow) – various calculation methods (Bermudez-Saldana and Cronin 2006; Knauer et al. 2007; Murawa et al. 2006; Wang et al. 2009; Yan et al.2006; Zvinavashe et al. 2009)

- Aqueous solubility (Murawa et al.2006)
- Chromatographic retention indices (Bermudez-Saldana et al. 2005)

Constitutive properties, i.e., the number of functional groups – important for heterogeneous data sets:

- Sub structural fragments (Casalegno et al. 2006)
- Different mathematical approaches on the numerical characterization of molecules with chiral center(s) (Natarajan and Basak et al. 2011)
- Presence of ether linkage, hydrogen bond donor groups and acetylenic carbons (Kar and Roy 2010)
- Heteroatom Corrected Extended Connectivity Randic index ((1)X(HCEC)) and the Density Randic index ((1)X(Den)) (Senior et al. 2011).
- Descriptors from the Dragon software: qH(+) (maximum net positive H atomic charge), spatial autocorrelation (MATS7m); n-O (count of oxygen atom); y component of dipole moment (Dip(y)); V-SA(2); globularity (Glo); indicator variable I-CH.(CH3) and R-PC
  (negative partial charge); MATS8P; Mor24u; MW; Mor14e; Mor20m MATS3v; HOMA (Mazzatorta et al. 2005; Porcelli et al. 2008; Yan et al. 2006; 2008)
- the vertex degrees ((EC)-E-0), the extended connectivity of first order ((EC)-E-1), and the numbers of paths of length two (P2) (Toropov and Benfenati 2006)

Atom-level, quantum chemical and topological parameters (related to factors including binding to the receptor site):

- The quantum chemical parameters obtained from ab initio calculations and optimization e.g. using the B3LYP/LANL2DZdp-ECP methodology (Senior et al 2011).
- Less negative charge surface area (Kar and Roy 2010)
- Energy of the lowest unoccupied molecular orbital, E(lumo), and the energy of the highest occupied molecular orbital, E(homo) (Zvinavashe et al. 2009)
- Difference between the E(homo) and E(lumo) (Mazzatorta et al. 2005)
- Dipole moment (Praba and Velmurugan 2007)
- Hydrogen bond donor and acceptor ability (Devillers 2004)
- Atomic charges / superdelocalisabilities on relevant atoms, e.g., P / O (Bermudez-Saldana and Cronin 2006)
- Amplitude range for affinity, e.g., K(a) and phosphorylation constants, K(p)(Mastrantonio et al. 2008).

Receptor modeling approaches (Classic drug design based on receptor ligand interactions)

- CoMFA (Slavov et al. 2008)
- the Catalyst programme for building pharmacophores (El Yazal et al. 2001)
- Computational 3-D pharmacophore models (El Yazal et al. 2001; Gupta and Mohan 2011)

- Comparative Molecular Similarity Index Analysis (CoMSIA) (Roy et al. 2008; Zhao et al. 2004)

The Panel noted that the choice of descriptors for modeling should be made with knowledge of the complexity of effect(s) being modeled and the mechanism of action. Therefore, constitutive descriptors may be important in a model for a diverse or chemically heterogeneous dataset. Within a congeneric series, descriptors relating to receptor binding (e.g. charges / superdelocalisabilities) will become important. For a very well refined dataset (i.e., one of high quality), 3-D approaches such as CoMFA may be appropriate. All of these approaches have been shown to be of use for modeling receptor binding related to toxicity endpoints, e.g., estrogen receptor binding (Shi et al. 2001; Tong et al. 1998).

One issue that should come to the fore in modeling is the relevance of the mechanism of action and / or AOP. Should a choke point be found, directed testing (e.g., of AChE inhibition) could be performed to help develop the domain of the AOP. This, in turn, may assist in the elucidation of appropriate descriptors.

Charge Question 2: ***Web-ICE for Predicting Acute Toxicity for Aquatic Animals***

The white paper investigated the use of Web-based Interspecies Correlation Estimation models (Web-ICE) to predict acute toxicity values and potentially fill data gaps (described in ***Appendix B)****. The review of the current version of Web-ICE suggests that, when using the minimum acute toxicity data for aquatic animals that are typically submitted to fulfill requirements under FIFRA (i.e., data for Daphnia magna, rainbow trout and bluegill sunfish), Web-ICE appears to generate the most precise predictions of toxicity for closely related fish species (e.g., within the same family). If toxicity data are available for either rainbow trout or bluegill sunfish, Web-ICE may be useful for predicting the toxicity of a chemical to some species of fish. The current review of this tool indicates that Web-ICE does not have sufficient data to fulfill the minimum data requirements across all taxa defined in the 1985 Guidelines when there are only acute surrogate data available for rainbow trout, bluegill and D. magna; however, this may change*

a) Appendix B of the white paper includes a review of the utility of Web-ICE for quantifying sensitive acute toxicity values for aquatic animals exposed to pesticides. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this review.

Panel Response

The Panel indicated that the Interspecies Correlation Estimation (ICE) is an excellent program to estimate the toxicity of a chemical to a species for which there are no measured values. Web-ICE is the interface for the program on the web. The program works by first setting up a correlation between two species for acute toxicity responses to any chemicals for which there are data. Once the correlation is established, then the response of one of the species to a new chemical can be used to estimate the response of the second species, without having to obtain empirical data. This correlation is currently independent of MOA.

The Panel considered the conclusions made by the EPA in Appendix B regarding the utility of Web-ICE to be reasonable. The Panel agreed that there were numerous strengths to this program especially in the potential ability to predict toxicity of a chemical to a species within the same taxonomic group for which large data gaps exist, for example for a threatened or endangered species. The fact that the program is transparent and freely available on the internet increases its utility. However, the program is limited by only incorporating acute toxicity responses and by not being applicable to predict toxicities across different phyla. In order to be useful to predict chronic toxicity, the program will have to incorporate mode of action of the chemicals, specifically geared to defining adverse outcome pathways that are linked to endpoints relevant to growth and reproduction. Additional recommendations to increase the utility of the program included adding estimates of confidence to the predictions, eliminating data redundancies that currently reduce functionality of the program, and potentially incorporating (Q)SAR approaches for specific chemicals and comparing species with similar life history strategies. The specific strengths and weaknesses of Web-ICE are enumerated below.

1) **Strengths of Web-ICE:**

- The concept and practice of interspecies relationships is well established and the three separate modules that have been included in the program (single species estimates, endangered species, and species sensitivity distributions) are an excellent resource for regulators and academic scientists.
- The Web-ICE software brings together a remarkable number of toxicity data; there are more than 3,000 ICE models available. It is an excellent way to use empirical data to fill in data gaps for other species.
- The Panel agreed that the closer the species, the better the ICE interpolation of toxicity. The greater the distance between taxa, the poorer the correlation. This makes sense as physiology, metabolism, etc., will be similar for closely related species and will vary for species that are further apart. However, it is likely that the correlation also depends on the type of chemical investigated and the MOA. For some chemicals (e.g., chemicals whose acute MOA is narcosis), it is likely that the prediction will work across a wide range of species, since most species will be sensitive to a narrow range of concentrations of the chemical. For other chemicals (e.g., endocrine disruptors), the predictions will probably only work for close taxonomic relationships. Thus, the MOA is likely to have an effect on the usability of the algorithm for wide predictions.
- This program was viewed as an excellent method to estimate toxicity of chemicals for threatened and endangered species. The Web-ICE program provides utility, specifically when there are surrogate species that are in the same taxonomic group and similar life history as the threatened/endangered species. The approach is limited currently, however by the availability of empirical data for non-standard organisms.
- The current exercise importantly explored the potential for filling in data gaps associated with regulatory MDRs as described in the 1985 Guidelines used by OW to establish aquatic life criteria.
- The approach is broadly applicable across species. There are models of varying quality for many relevant species.

- The statistical method applied, Model II regression, is appropriate for the development of models where there is error in both the dependent and independent data.
- The general rules found in Appendix B, which are common for most statistical tests (low mean square error, cross-validation, high degrees of freedom, high R^2 value, low p-value, narrow confidence intervals), as well as close taxonomic distance for comparison species, are a good place to start to evaluate the ICE. However, there are several issues that will need to be addressed to be able to assess the overall quality, as described below under weaknesses.

2) Weaknesses of Web-ICE:

- As noted in Appendix B, data redundancies and errors in Web-ICE require attention. Nominal vs. measured data should specifically be examined for data quality and perhaps nominal data should be flagged so that its use in extrapolations is transparent. It might be interesting, when possible, to allow the user to include or not include the flagged nominal data.
- The overall quality of ICE values rely on a number of issues, for example, on the number and spread of data (i.e., a distribution of toxicity values over five orders of magnitude will almost certainly have a better correlation than one over 4 orders of magnitude, etc.).
- The Panel recommended the inclusion of MOA, especially employing AOP conceptual approaches, as a further refinement of Web-ICE to determine acute toxicity relationships.
- There is an implication (e.g., point 4 in the final paragraph, p 4, Appendix A) that predicted toxicity values (ECOSAR) are used to form the ICE. If this is correct, great caution must be exercised as predicted values will be less accurate, and it may be more accurate to predict the toxicity value directly. Predicted toxicity values should also be flagged.
- Domains of applicability could be provided, as with some estimates of confidence in the prediction.
- Since ICE correlations are stronger within mechanisms of action, for application to pesticides, it may be better to separate chemicals out into different mechanisms, if possible. If the predictions from an ICE will be used for regulatory purposes, a reporting format (which could be automated) may be required. The (Q)SAR Model Reporting Format (QMRF) is a good example of a template that could be adapted for this purpose and simplified. Likewise, perhaps guidance should be provided on when to accept (or have high confidence in) an estimation from an ICE. For instance, the OECD guidance document, *OECD Principles for the Validation of (Q)SARs* (OECD 2007), has been very successful in promoting better development and application of *in silico* models and key elements, e.g., transparency, use of an unambiguous algorithm.
- The inclusion of both a fish and *Daphnia* species to increase predictability across all species probably depends on the MOA. This may be true for narcotics, but may not be true for chemicals that act through other MOAs or that operate via distinct MOAs in invertebrates vs. vertebrates, e.g., endocrine disruptor chemicals often act

differently across different taxa. The EPA should consider building models for specific MOAs.

- When deciding that organisms are similar, their life histories should be included, as there could be very different conclusions if different sensitive windows of susceptibility are tested. For example, in fish that spawn once a year, it could be hard to extrapolate toxicity values obtained for one species during gonadal recrudescense with another species during prime reproduction, as these two different time points may exhibit different sensitivities. This will be particularly true if chronic exposures are included. Assessing overall quality of predictions will rely on a number of issues, not least on the number of data sets available and the variability of the data (i.e., a distribution of toxicity values over five orders of magnitude will have a better correlation than one of four orders of magnitude, etc.). As it is currently configured, the SSD module for ICE has applicability to general toxicity (narcosis). This module may not work with pesticides that have complex MOAs.

b) Please comment on potential modifications to the existing version of Web ICE that could improve its ability to predict the acute toxicity values of pesticides to different species of aquatic animals. For example, expansion of the tool to consider MOA, addition of data for infrequently tested species that may fulfill the data requirements defined in the 1985 guidelines.

The Panel had a number of suggestions to improve the existing version of Web-ICE, version 3.1. Most of the suggestions applied to improving the prediction of acute toxicity, although some improvements were suggested that included predictions of chronic toxicity among species. The major recommendations are discussed below.

1) Systematic evaluation of each of the MDRs to determine if every species listed is recommended for every chemical/MOA and system.

The EPA stated that Web-ICE was not sufficiently predictive to fulfill the MDRs; therefore, it could not be used in this manner. However, the Panel felt that this was primarily due to a lack of data in ICE for non-chordate and non-arthropod species (i.e., MDR #7 in the 1985 Guidelines or simply, MDR7). For some chemicals, the MDR7 requirement may not add much information to the determination of Final Acute Values (FAVs) or of HC5s. If the MDR7 does not add value for a particular chemical, perhaps it should not be included in these determinations. For example, species tested to meet MDR7 do not have impact on endpoint selection for AChE inhibitors and narcotics (see Figure 3 and Table 5, Appendix D, p 12). Species tested to meet MDR7 are the least sensitive to AChE inhibitors and do not differ in sensitivity to narcotics as compared to other MDRs. This means that MDR7 provided very little scientific value in the determination of FAVs or of HC5s.

The Panel recommended a systematic sensitivity analysis be conducted to determine the extent to which individual MDRs contribute to the endpoint estimation in general, as well as for each MOA. The Panel noted that data exist in the current set of analyses in the white paper and appendices, e.g., the SSD analysis (Appendix C) and EF analysis (Appendix D), to test the value-added of each species needed to fulfill a MDR and to see if they separately, or in combination,

have import on the determination of endpoints for many MOAs. The Panel suggested that additional studies on MDR7 species would be useful to fill this data gap for certain MOAs.

2) **Inclusion of chronic effects, particularly those related to growth and reproduction, as well as consideration of life history is recommended.**

As currently constructed, Web-ICE relies only on acute exposures and lethality. A major advance for this program would be to include sublethal endpoints from chronic studies. These should be framed within AOPs that link to reproduction, growth and survival. Future ICE correlations that consider chronic endpoints related to growth and reproduction should be constructed within species of similar life history strategies. At the population level, organisms have evolved growth and reproductive patterns/tradeoffs that are defined by environmental constraints on survival. These growth and reproduction patterns can vary widely within taxa (for example, short life span, fast growth, high reproductive effort versus long life span, slow growth and reproductive effort spread over several years). As such, chronic effects of contaminants can have very different effects on species with different life histories or reproductive strategies (Spromberg and Birge, 2005).

3) **More directed data collection to fill in points along the MOA and adverse outcome pathways, particularly for those chemicals that have specific effects is recommended.**

- Adverse outcome pathways will probably influence the use of ICE across species that are in different taxa if chronic data can be included. It would be good to work towards being able to predict across taxa. Thus, integrating MOAs will help to improve predictability of the algorithm.
- ICEs have been found to work optimally within defined mechanisms of action. Thus, the extrapolation of the effects of non-polar narcotics can be very accurate. As toxicity mechanisms become more "specific", the ICE may (this would have to be investigated) become weaker. It may be that ICE for specifically acting pesticides are improved by restricting them to MOA. There may be further parameters (allometric, metabolic, physiological or otherwise) that could be included to explain the inter species differences and hence improve the correlation.
- It was not clear from the discussion which MOAs would be included in the Web-ICE; however, this choice should be very flexible and dynamic to eventually include molecular endpoints that are correlated to population level effects, such as decreased vitellogenin in females (Miller et al 2007) and plasma hormones (Ankley et al., 2008; Murphy et al, 2009).
- The current SSD module for ICE is configured for a general mode of toxicity (narcosis) and would need to be upgraded to include pesticides that have complex modes of action.

4) **Inclusion of ICE modules for algal and plant species is recommended.**

The Panel also suggested that the model output from Web-ICE would be better understood if the data were provided in a more user-friendly fashion, e.g., graphing the output with the goodness

of fit displayed or using a model averaging routine and downloading the output into a spreadsheet program.

The Panel noted that there should also be more access to data and outputs from SSD models as well as good graphical display of the SSDs.

Charge Question 3: ***Use of SSDs to Estimate HC5 using Varying Amounts of Data***

As described in ***Appendix C****, USEPA investigated the use of species sensitivity distributions (SSDs) with varying amounts of empirical data to estimate the fifth percentiles of sensitivity distributions of toxicity data (termed the "HC5"). Analyses performed by USEPA for this effort suggest that SSDs have limited utility when test data points are limited, specifically where biological diversity is not sufficient to represent the distribution of taxa sensitivity. The analyses further suggest that SSDs applied to datasets that satisfy existing taxonomic requirements with extrapolation constants (as defined in* ***Appendix C****) would likely result in lower and yet reasonable approximations of the HC5.*

a) Appendix C of the white paper includes a description of the analyses of SSDs. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this analysis.

Panel Response

The Panel concluded that EPA's analysis described in Appendix C represents a logical and well thought out consideration of the adequacy of available data, MDRs, and methods for estimation of the HC5 using SSDs. The resampling and distributional approaches to estimating HC5 values are based on standard statistical analysis tools and estimation approaches; therefore, they do not present any new or novel methodological issues. Maximum likelihood and moment estimators have been studied extensively, at least as they are used to estimate the first (mean) and second (standard deviation) moments of a set of data from which a random sample has been drawn. The literature on the adequacy of the graphical approaches to estimation of these distributional parameters is much less. Estimation of specific percentiles of these distributions has also received extensive review by statisticians. In general, these approaches are based on empirical relationships derived from statistical analyses and not on a mechanistic or biological model. Without greater mechanistic or biological understanding it is not possible to explain why a specific combination of assumed distributional form, estimation method, sample size, and sample constraint is unable to produce acceptable estimates of the HC5 for a particular situation. The strength and weaknesses of the resampling approach vs. the distributional approach are discussed below.

1) Strengths and weaknesses of the methods

The Panel thought that both the resampling approach and the distribution simulation approach were equally effective in demonstrating that:

- Decisions regarding (i) the distributional form for the SSD, (ii) the sample size, and (iii) the method of sampling, will affect dramatically the estimated value of the HC5.
- Demonstrating that both approaches produce HC5 estimates that are biased (either under- or over-estimate the true HC5 value).
- Demonstrating that increasing the sample size will simultaneously decrease both the bias and the uncertainty (standard error) of the HC5 estimate.
- Randomly sampling from a theoretical SSD (the distribution simulation approach) or from a large set of LC50s (the resampling approach) allows estimation of one-sided confidence intervals with coverage levels close to theoretical expectations, e.g., estimated 95% lower confidence bounds actually have close to 95% coverage.

The distributional approach alone demonstrates that:

- Traditional distribution goodness-of-fit statistics and the Akaike Information Criterion (AIC) – a typical measure of goodness of fit used to compare model fits of very different form) index alone or in combination are inadequate for choosing the best distributional form to describe a particular dataset.

The resampling approach alone demonstrates that:

- Using sampling with replacement from a set of data will produce 95% lower confidence intervals that do not reach the expected 95% coverage when small sample sizes are used.
- Using sampling constrained by the MDRs or FIFRA data requirements can, in many cases produce 95% confidence intervals that are almost guaranteed to contain the true HC5 but their actual coverage may be much greater that 95%.

These findings are not new and would have been predicted from the statistical properties of the estimation methods used, from the characteristics of the distributional forms examined and from statistical research that illustrates how inadequate goodness-of-fit statistics are when applied in situations with small sample sizes. Statisticians know that the ability to make decisions or estimate parameters with any degree of certainty depends critically on having adequate numbers of data points.

The two approaches complement each other. The strengths of the distributional simulation approach, i.e., ability to examine what-if questions about the impact of distributional shape on effectiveness of parameter estimation, biases in estimation of the HC5, and effectiveness of goodness-of-fit statistics point to the true distributional form, are weaknesses of the resampling approach. Similarly, the strengths of the resampling approach, i.e., ability to explore statistical properties using real data without having to specify a distributional form and ability to answer what-if questions regarding the effect of extreme data points, are weaknesses of the distributional simulation approach.

2) Strengths and weaknesses of the results

The Panel found that the results presented in Appendix C were clearly and logically presented, although they noted several weaknesses in the analysis. Application of multiple estimation

methods was the major strength of the analysis. The Panel recommended that the graphical method be used for all of the distributions, not just the normal, logistic and triangular distributions because this method is based on quantiles of the fitted distribution.

The Agency used five different distributions to both simulate data and fit distributions to these data. The five were the log-normal, log-logistic, log-triangular, log-symmetric, and Pareto distributions. The normal, logistic and triangular distributions are based on quantiles of the fitted distribution. These three particular distributions have “closed” form estimates; therefore, there are simple equations that allow parameter estimation directly from the raw data through a set of simple computations. The other distributions do not have “standard” or “reference” forms such as the "standard normal” form and there are no simple “closed form” equations for computing the parameter estimates. In this situation, computing parameter estimates requires iteratively solving a set of non-linear equations. Some parts of these non-linear equations are themselves non-linear functions that further complicate the computations. The process is computationally difficult, but not impossible.

Maximum likelihood methods for the normal, logistic and triangular on the log scale are similar to least squares methods in that the method tends to weight more heavily the extremes than the middle of the distribution. In addition, for these three distributions, the method of moments produces results similar to MLE. The graphical approach on the other hand tends to weight the middle of the distribution more heavily.

The graphical estimation method performed better than the maximum likelihood or method of moments based on current statistical understanding of these methods. Parameter estimates for the graphical method only require three points to fit a line from which intercept and slope values are obtained; hence, the method has greater applicability for small sample size situations.

One large benefit of the graphical method is that the Shipiro-Wilks goodness-of-fit statistic, based on the R^2 of the line fit, is readily available as a by-product of the fitting method. A high R^2 value tells us that the data are likely to come from the assumed distribution. During one presentation, a public commenter (for CropLife America) indicated a preference for the MLE approach and Anderson-Darling goodness-of-fit test. The Panel noted that MLE is available in some form for all the distributions considered, but the same can be said for the graphical methods. The statistical properties of MLE estimators have been extensively studied, but many of these are large sample or asymptotic properties and are not very relevant in small sample estimation situations. Similarly, the Anderson-Darling goodness-of-fit test is one that integrates the area between the empirical distribution and the fitted distribution and as such emphasizes a more uniform goodness-of-fit. The Shipiro-Wilks goodness-of-fit test is available for small sample sizes and has easy description. While it may have poor properties for small sample sizes, the same can be said for all distributional goodness-of-fit statistics and in this case at least critical values are available.

Figure 1 (p 11 and Table 1, p 10 in Appendix C) displays the cumulative distribution functions (CDFs) and probability density functions (PDFs) for the specific forms. The Panel inspected the PDFs and CDFs of the specific distribution forms and concluded that the PDFs of the specific

distributions have roughly the same 80th percentile and the CDFs of the specific distributions are designed to have a wide range of theoretical HC5 (5th percentile) values.

The Panel identified several weaknesses in the analysis:

Lack of justification for the specific distributional forms. The EPA did not provide a good discussion of why the specific distributional forms used for the distribution simulation approach were chosen. With such a justification, it may have been possible for the EPA to extend these results to an assessment of how different distributional shapes affect the accuracy and precision of HC5 estimates.

On the log scale, the log-normal, log-logistic and log-triangular distributions were symmetric about their respective means or mode in the case of the log-triangular distribution. On the untransformed (anti-log) scale, these three distributions are negatively skewed with the bulk of their density shifted toward the lower concentrations with a low probability for very large concentrations in right tail. Two of the distributions, the log-normal and log-logistic, are typical theoretical forms that have wide applicability in analysis of concentration data. The symmetric log-triangular distribution represents a negatively kurtosed distribution (on the transformed scale), which means that it has lower peak density and fatter tails than would be expected from a normal distribution. It is a model for data from a sampling situation where the sampling was targeted to ensure uniformity across a broad spectrum of concentrations. The Wiebull and Gumbel distributions are also negatively skewed distributions. The parameters chosen for the Weibull make it more negatively skewed than the selected Gumbel. The Burr (generalized log-logistic) distribution is described in one place as having parameters *a;b;c* (Table 1, p 10, Appendix C) and in another as having *b;c;k* (p 26, Appendix C) so it was difficult to determine its exact shape. It appeared to be more negatively skewed than either the Weibull or Gumbel. The Pareto distribution is a member of the power family of distribution functions and it has statistical properties that make it very difficult to work with. For the parameters chosen it was very negatively skewed.

Distributions did not have the same 5th and 50th percentiles. Another weakness of the analysis was the fact that the distributions were not chosen to have roughly the same 5th and 50th percentiles which would have facilitated comparisons of HC5 values across distributions with different "shape." In this study, both distribution location and scale were allowed to vary making it difficult to determine location or shape differences that result in a single distribution, the log-triangle, to produce good results and another, the Pareto, to produce very poor results.

Considerations for censored data. One situation not discussed in the report was what to do with censored data that do not allow direct estimation of the LC_{50}/EC_{50}. Graphical methods can be easily adapted to handle this situation. One of the public presenters (for CropLife America) mentioned that MLE estimation can be used in the censored data case, but computation becomes much more difficult. The Panel commented that handling censored data with moment estimators is even more complex for the simple normal case and near impossible for any other distribution. Additionally, the public commenter emphasized the importance of setting a standard for the plotting position of the empirical distribution

function computation, i.e., the *y*-axis values. The Panel agreed with this statement and stated that the plotting position does have an effect on the estimation of the HC5 and hence one particular protocol should be used.

Situations where underlying data come from a mixture of distributions. Another weakness of the study was that it did not explore situations where the true underlying data come from a mixture of distributions. The Panel noted that underlying data used to estimate a SSD may truly come from two or more underlying distributions. In this case, the empirical SSD represents a mixture of "conditions." In this situation, no one theoretical distribution regardless of the degree of skewness or kurtosis would fit the available data adequately. Normal *q-q* plots for some of the model fits to empirical data display the "hockey stick" appearance typically observed when attempting to fit a single theoretical distribution to mixture data, rather than the convex or concave appearance typically observed when fitting a wrongly shaped theoretical distribution to a set of data. Analysts need to be aware of this and carefully question the data to understand if in fact there are two or more "types" of data, for example, the situation where say salt water species were less sensitive (larger average LC_{50} values) but also more variable in their LC_{50} values than fresh water species.

Estimation of uncertainty. One of the public presenters (on behalf of CropLife America) indicated that the uncertainty related to estimation of HC5 is much larger than that related to HC10 which is less than that for HC50. The Panel stated that this is a well-known characteristic of estimation of extreme events. The uncertainty is a function of the percentile being computed and is roughly related to the value *p(1-p)* for estimation of the p^{th} percentile. Another well-known characteristic is that uncertainty in the estimate is also going to be proportional to the square root of the sample size. From this, the uncertainty based on a sample of size of n=9 versus n=81 will only differ by a factor of 3 which might be swamped by the *p(1-p)* component of the uncertainty.

3) Strengths and weaknesses of the conclusions

All five distributions (log-normal, log-logistic, log-triangular, log-symmetric, and Pareto) were negatively skewed to some extent; however, "skewness" in the log-normal, log-logistic and log-triangular distributions does not have to be accounted for in the fitting process because these distributions can be transformed to symmetry. Skewness and variance are indistinguishable with less available data and as sample sizes decrease, it becomes increasingly difficult to estimate variance independently from skewness.

The Panel noted that the study conclusion indicating that the log-triangular distribution produced the best results across all AChE inhibitors was primarily due to the resampling study results. The underlying data for this resampling study is likely to be less random and more uniform in its representation of the range of concentrations in populations across taxa. The Panel conjectured that these factors make the log-triangular distribution a favored choice among the distributional forms. In addition, the use of GMAVs would further lead to a more uniform representation and hence favor the log-triangular distribution. From the distribution simulation study it was observed that when the data are sampled from a (rue or known log-symmetric distribution, the distributions in which the skewness cannot be removed by a log transformation (that is, not the

log-normal, log-logistic or log-triangular distributions) are at a disadvantage. These non-transformable symmetric distributions have mathematical forms that are not flexible enough to fit data from a truly symmetric distribution. The Pareto distribution is particularly hampered because in addition to being skewed, its mathematical form is more rigid than the others.

The Panel indicated that data quality assurance and validation of all values used in the analysis were a necessary part of the assessment to determine the utility and representativeness of the fitted SSDs. Additionally, uncertainty in all its forms, i.e., relevance to the real world, human errors, random variation, and parameter uncertainty, should be explicitly and transparently documented. One Panel member suggested that the geometric mean not be used in situations where endpoints vary by more than a factor of 10 for data from the same species. There can be wide differences within the same species tested in different laboratories. The Panel noted that the EPA was aware of these issues and that additional efforts are planned to continue to assess and validate the data used in the analysis.

The Panel stated that the conclusions were clearly articulated and supported by the results except for one: "In general, distributions on transformed data performed better than distributions on untransformed data." The Panel noted that it would be premature to conclude that transformation was the key factor in better performance because the effect of transformation was difficult to extract from distributional shape. Because it is difficult to extract the effect of transformation from distributional shape, it is difficult to pick on transformation as the key factor in better performance. All of the distributions used were negatively skewed to some extent; the difference was that for the lognormal, log-logistic and log-triangular distributions, skewness did not have to be accounted for in the fitting because they could be transformed to symmetry. With less data, skewness and variance are indistinguishable and hence it becomes increasingly difficult to estimate one independently from the other.

The Panel recommended that future assessments should be based on data that support curve fitting rather than estimates such as NOECs or LOECs that have limited utility in endpoint estimation (see for example, Bailer and Oris 1999; Warne and Van Dam 2008; Landis and Chapman 2011; Jaeger 2012). Ideally, the future risk assessments would be based on high quality and validated quantitative data capable of providing accurate EC*x* estimates. Studies where these data could be collected should be encouraged as standard practice rather than those studies that have provided NOECs or LOECs. Historical NOECs and LOECs would be used when these are the only data available and best professional judgment determines they provide useful information for the risk assessment. In this case, a statistic such as a MATC (maximum allowable toxicant concentration the geometric mean of the NOEC and LOEC) should be used. As mentioned by both the EPA and public commenters representing CropLife America, a variety of curve-fitting and goodness-of-fit approaches should be applied to this problem to ensure that the right set of tools are used for the right situation.

The Panel indicated that the section on model averaging was the most difficult to understand and hence had conclusions that were the most difficult to assess. One Panel member questioned why anyone would even think of including the rationale for inclusion of the Pareto distribution into the mix of models to be averaged. With small sample size, the AIC values for all the models were large; hence, all the models seemed to be equally weighted in the mix. An alternative

weighting parameter might be the R^2 statistic from the graphical goodness-of-fit method. Giving higher weight to more linear fits would reward distributions that are better able to conform to the data and penalize distributions that do not conform. As long as the distributional forms have two or three parameters, an extra parameter should not make much difference in the weighting.
The conclusions regarding the effect of restricting the selection of taxa to conform to MDR or FIFRA data requirements were also expected from a statistical point of view. The intent of the requirements was to select taxa that span the range of sensitivities to the chemicals under study. Such selection tends to produce samples that are more "uniform" than would be expected by simple random sampling. This kind of selection would support the choice of the log-triangular distribution and the use of graphical methods for the reasons discussed previously. However, if the full AChE inhibition data set were used and the empirical distribution function was plotted against the theoretically cumulative distribution function of the fitted SSD, then the empirical to fitted quantile-quantile plot would be created. If this plot is very linear, indicating a good fit of empirical to theoretical model, then the resulting HC5 estimate should be quite accurate. If the *q-q* plot has a convex form, indicating that the model does not quite fit as well, then the estimated parameters from the graphical method will result in an HC5 estimate that is an underestimate of the true value. If the *q-q* plot has a concave form, then the HC5 estimate will be an overestimate of the true value. Restrictions on sampling effectively specify that a point or two are required from different "thirds" of the concentration distribution. Negative distributional skewness with correspondingly low but positive chance of some large (right tail) values ensures a convex *q-q* form and hence ensures underestimation of the HC5 value.

4) Additional comments on Appendix C

The Panel had the following additional comments on Appendix C:

- Section 2.2 reviewing basic statistics related to SSD distributional form and parameter estimation was a positive addition to the discussion. The discussion of the effects of transformation of data on the comparison of distributional fits (Section 2.2.6) and on the estimation of bias (Section 2.2.7) was well written. On the other hand, Section 2.2.10 on how sampling variance was estimated was totally inadequate, a mere two lines of text.
- At least one Panel member was confused by the use of the term "test results." In common language, a test might be an assay performed at a particular concentration for a given taxon and chemical. In this report, a test was actually a set of bioassay results that produced an LC/EC50 estimate. Using example data from Table 2 in Appendix C, it may have been easier to simply say that there were 171 bioassay studies of malathion on freshwater animals that produced LC/EC50 estimates for 71 unique species (for which SMAVs were computed) and 51 unique genera (for which GMAVs were computed). It might have been useful to know how many other bioassay studies of malathion were performed that were actually unable to produce an acceptable LC/EC50.
- There are two section 4.5's in Appendix C.

b) In Appendix C, several potential avenues for future work on SSDs were described, including i) extension of methods for ordered data to all continuous distributions ii) development of stepwise methods, including decision points, for data analysis to incorporate information on the relationship of MOA and taxa sensitivity, and iii) the use

of random effects methods for handling cases in which more than one test result is available for a given taxon. Please comment upon additional analyses relative to SSDs that could be considered by USEPA that may result in improved confidence in estimates of HC5 values using various amounts of data that are typically available for registered pesticides.

Panel Response

The Panel agreed that the SSD procedures described in Appendix C were useful in estimating the HC5. The precision of the HC5 estimates was improved when more of the MDR categories were available compared to less. The Panel commented on several areas of the analysis.

1) **Move away from a chemical-by-chemical approach**

The Panel stated that because small sample sizes were available for many individual chemicals, there would be large gains in the precision of the HC5 estimates by moving away from a chemical-by-chemical approach to one based on classes of chemicals sharing common MOAs (e.g., chemical and biological read-across or chemical borrowing). However, guidelines for developing robust chemical borrowing require additional development. Most of the Panel liked and supported the suggestion of using phylogenetic structure as a covariate or explanatory variable in some form of hierarchical or Bayesian modeling analysis. (Q)SAR modeling (see discussion to Charge Question 1) and IATAs (integrated approaches to testing and assessment) are introducing new computational and molecular tools to identify characteristics of chemicals that allow them to be grouped for assessment and which might also be used as covariates in subsequent analyses (see discussion to Charge Question 1). Used with hierarchical mixed effects modeling, this would allow more data to be brought into the assessment, increasing the confidence in resulting SSDs and HC5 estimates and supporting rapid assessment of broad classes of chemicals.

2) **Order statistics approaches may not be productive**

An order statistics approach to SSD parameter estimation via "graphical" methods may represent a less productive avenue of study. Sample size is the primary limitation to parameter estimation. Order statistic approaches are not necessarily immune to sample size issues. Order statistics are typically called on to address the need for more robust distributional forms than to address inadequate sample size. Once sample sizes get to the point that order statistic approaches are effective (say n >50), more formal distributional model fitting and goodness-of-fit actually tend to also be effective.

3) **Bayesian approaches are likely to be productive**

The Bayesian approach to parameter estimation expands the random effects model formulation discussed above by allowing specification of prior distributions on these parameters. As indicated in the white paper, a Bayesian approach introduces complexity to the interpretation of the findings and complexity to the interpretation of the estimated HC5. For example, the 95%

confidence interval becomes under a Bayesian framework a "95% credible interval", a quantity that many non-statisticians cannot readily describe.

A Bayesian approach should include consideration of the question of taxonomic representativeness. This was an issue identified by CropLife America as one requiring further research. Taxonomic representation and endpoints should be appropriate to the chemical being tested (i.e., based on the MOA).

Available data are used to specify/estimate a prior distribution that describes how a test species responds to a number of chemicals having similar MOA. This distribution specifies in a sense, what might be expected if one were to use this species in a study of a new chemical having the same MOA. If we assume that new data are collected for this species for a new chemical having the same MOA then this new data is combined with the (previously fitted) prior distribution to produce an estimate of the parameter needed, e.g., the HC5. This empirical Bayes HC5 estimate will have better statistical properties than would the HC5 estimate that is just based on the new data. The role of the prior distribution is to ensure that the HC5 estimator for the new chemical is not "too far" from what would be expected given the prior distribution. If, by chance, the new data indicated sensitivity that was much, much higher than say anything previously seen, the empirical Bayes approach would produce an HC5 estimate that was between the HC5 estimate predicted from the prior distribution alone and the HC5 estimate computed from the new data alone. In this way we say the empirical Bayes approach "shrinks" the new data estimate toward the center of the prior distribution.

4) Data and chemical characteristics may drive method successes

There was concern that the success of the analysis with the AChE data may be primarily due to the extensive and robust dataset available for this chemical. EPA is encouraged to expand the analysis to other chemicals and especially to chemicals with different MOAs, with potentially multiple MOAs and less robust data.

The Panel stressed that the EPA continue to look at the available data in assessing the appropriate level of protection, since in some cases it may be appropriate to use the lowest value to ensure a conservative level of protection. Professional judgment and broader understanding of the chemical under study and its MOA are needed when attempting to understand the effect of not having specific data for some species/taxa needed to fulfill a MDR. The Panel commented that the data to fulfill an MDR should support professional assessment and not supplant it.

5) Need to assess adequacy of subsets of MDRs

Both the resampling and distribution studies were performed under the assumption that all eight of the taxa classes defined under the OW MDR are available or necessary to properly understand the SSD and estimate the HC5. It is clear from the MDR class specific PDFs shown in Figure 3, Appendix D, that there are only two distinct types of distributions for the narcosis MOA, which suggests that only two or three MDR classes would be needed to characterize the distribution of acute responses. For the AChE MOA, only four to five MDR classes would be needed. The Panel expressed the desire to see a more formal, systematic assessment of the value of each

MDR in supporting estimation of the HC5 for example. Combinations of MDRs should be examined to determine which are most useful in estimating the HC5.

6) Utilization of specific species and most sensitive species testing

The Panel discussed whether there should be specific species testing to fulfill a MDR. Some of the Panel agreed with the use of specific species and others expressed reservations with this recommendation and were more comfortable with genera level MDRs. The Panel commented that comparisons of EEC values are typically performed with the most sensitive species. Similarly, species of special sensitivity to a specific MOA should be identified and used in the analysis to increase the certainty that the HC5 would be protective of all included species. The Panel agreed that testing across the life cycle of an organism should be continued to determine the impact of chemical exposure on each life stage.

The Panel examined the sensitivity of freshwater animals to different AChE inhibitors (Appendix C). They noticed that estimates of the HC5 were significantly improved when toxicity data from *Ceriodaphnia dubia,* a test species required by OW, is included with the toxicity data from the three preferred freshwater test species, e.g., *D. magna*, required by OPP to support a pesticide registration. The Panel referred to the whole effluent toxicity (WET) test methods used by OW to measure the toxicity of effluents and receiving water to freshwater, marine, and estuarine organisms (http://water.epa.gov/scitech/methods/cwa/wet/). The WET test methods use *C. dubia* as the preferred freshwater invertebrate test species. They recommended that OPP investigate how inclusion of *C. dubia* toxicity data (in addition to the preferred freshwater species) affects choice of distributional form and ability to estimate parameters of the chosen SSD. In addition, using the AChE-inhibitor datasets, the Panel wondered how the HC5 calculated using *C. dubia* data compares to the value calculated using *D. magna* data. The Panel suggested that the EPA delve further into this issue by exploring differences in HC5 estimations when different species-specific toxicity test protocols are available.

The Panel recommended that EPA continue to evaluate SSD procedures for pesticides with different MOAs and investigate the differences in SSDs and HC5 estimations using species-specific protocol test results. For example, as was the case for observing differences in sensitivities between *C. dubia* and *D. magna* to AChE inhibitors, there are also differences in sensitivity between Cladocera and amphipods to pyrethroid pesticides. The amphipod *Hyalella azteca*, required as part of OW's MDRs, has been shown to be particularly sensitive to pyrethroid pesticides, and is considerably more sensitive to this class of pesticides than *D. magna*. The Panel recommended that EPA investigate how inclusion of data using *H. azteca* affects SSD distributions with pyrethroid pesticide data and compare how the HC5 calculation with the inclusion of *H. azteca* data compares to results using *D. magna*.

7) Use of exposure and toxicity modifying factors

The Panel suggested that exposure and toxicity modifying factors (ETMFs) should be considered to enable data collected under controlled laboratory conditions to be translated to realistic environmental conditions. For many chemicals, peak concentrations in nature typically coincide with changes in major water characteristics, such as salinity, pH, hardness, etc., can be monitored

quite easily, e.g., coastal marine waters. Aquatic exposure models such as PRZM/EXAMS can be used to predict typical levels of these characteristics at critical events. Studies that generate the data used to estimate the HC5 should be conducted, where possible, with ETMFs set at or near values known to produce peak concentrations in nature.

8) Use of chronic endpoints

The Panel indicated that chronic endpoint data collection, methodologies and availability are not at the same level as they are for acute endpoints data. This situation will change as EPA increases use of AOPs and when data from the endocrine-disruptor screening program (EDSP) become available. Once available, these data will support the development of robust SSDs for various MOAs.

Charge Question 4: ***Extrapolation Factors for Estimating Acute HC5s with a Specified Level of Confidence***

The EPA OW's aquatic life criteria (ALC) involve calculation of the "Final Acute Value" (FAV) as the fifth percentile of acute toxicity values for genera of aquatic animals, and impose certain minimum data requirements (MDRs) for such calculations. To provide some assessment of effects when insufficient data are available for calculating an FAV, EPA's Great Lakes Water Quality Guidance (GLWQG) specified extrapolation factors (EFs) by which the acute toxicity value for the most sensitive tested aquatic animal genus could be divided to provide a conservative estimate for what the FAV should be if sufficient data was available (EPA 1995). In this case, conservative estimates of the FAV are generated by accounting for uncertainty in possible EF values, which generally results in an underestimate of the FAV that may be generated using a data set that would be considered complete according to the 1985 Guidelines. The EFs incorporated into the GLWQG were based on analyses by Host et al. (1995) of toxicity test datasets used in freshwater ALC that existed at that time; these analyses involved repeatedly subsampling (resampling) of these datasets to assess the relationship of the lowest toxicity values in datasets of a specified size and composition to FAVs based on larger datasets.

The analyses described in ***Appendix D*** *of this white paper builds upon the analysis conducted by Host et al. (1995) by (a) utilizing more recent aquatic toxicity data and data for chemicals that do not have an ALC, (b) developing EFs for different MOAs to reduce uncertainty, (c) paying specific attention to EFs for OPP minimum data requirements for acute toxicity testing of aquatic animals, and (d) generating EFs using statistical distributions descriptive of acute toxicity values for different taxonomic groups and based on multiple chemicals, rather than resampling toxicity data for individual chemicals. These analyses established the feasibility of establishing EFs for certain MOAs, demonstrated how a distributional rather than a resampling approach can more effectively exploit data to develop MOA-specific EFs, and documented the relationship between OPP and OW benchmarks. In addition, the current efforts considered how EF applicability and meaning might be affected, relative to the GLWQG, by different strategies for assigning reference HC5 and for summarizing EF distributions.*

(a) *Please comment on the extent to which consideration of MOA decreases uncertainty associated with HC5 values that are estimated using EFs.*

Panel Response

Background

The EPA developed EFs to provide some assessment of effects when insufficient data are available for calculating a FAV. While SSDs are useful for large data sets, EFs are useful for small data set estimations. Conservative estimates of the FAV are computed by accounting for the uncertainty in possible EF values, which generally results in an underestimate of the FAV that may be generated using a data set that would be considered complete according to the 1985 Guidelines. Classification of chemicals with similar MOAs or AOPs provides an opportunity for testing this concept of EFs by lumping toxicity data for different species and different chemicals with similar MOAs. EPA presented data testing these concepts in Appendix D.

General Conclusions and Recommendations

EPA tested the validity of the MOA approach using AChE inhibitors and narcotic agents (Appendix D). The Panel concluded that these examples indicated the use of the MOA approach may decrease uncertainty associated with HC5 values, especially those estimated using EFs. The MOA approach has a strong logical foundation for AChE inhibitors. AChE inhibitors binding to the same target (AChE) will elicit similar effects (increasing inhibition with increased dose for each chemical) but with differing toxicity thresholds and ranges for each inhibitor.

Results showed that there were positive aspects of using MOA and AOP approaches in these analyses. First and foremost, it is somewhat intuitive that compounds, such as AChE inhibitors, will exhibit varied receptor binding affinities among chemicals and that different taxa will have varying sensitivities. The Panel recognized that invertebrates (e.g., arthropods) were more sensitive than fish to the acute effects of AChE inhibitors. This was the case for all six AChE inhibitors tested. The Panel encouraged EPA to further develop the MOA approach for chemicals with other MOAs, e.g., synthetic pyrethroid insecticides. Evaluation of sets of chemicals with different MOAs will provide a more robust database in which to evaluate uncertainty in estimated HC5s. The Panel cautioned that close attention should be paid to the most sensitive species which may differ for each chemical with the same MOA. They also stressed that factors such as the relevance of resident versus non-resident species, sensitivity of individual species, intra-species sensitivity, protection of endangered or threatened species, and use of information gained from studying the MOA of AChE inhibitors are important considerations to guide future ecological risk assessment of pesticides with different MOAs. The Panel strongly recommended the Agency evaluate the importance of each species to meet a MDR. That is, some species (acute toxicity values) may have little added value (if any) in estimating the HC5. In many cases, the Panel commented that fewer species might be needed to meet a MDR depending on the MOA.

Aquatic invertebrates were noted to be the most sensitive species to AChE inhibitors. Perhaps, this information can be used to guide future research and monitoring efforts to confirm the sensitivity of aquatic invertebrates in the aquatic community. For example, mesocosm studies might serve as a first tier confirmation route and field monitoring studies would serve as a second tier confirmation route. Such approaches might be useful to other MOAs and AOPs.

Finding consistent effects throughout this continuum of ecological assessment approaches is an important factor in protecting aquatic communities.

The Panel provided the following general recommendations:

1) **EPA should continue to develop this AOP approach using EFs and HC5 methodologies.**

   Development of more AOPs will enhance our knowledge of the value of these approaches and by helping us to gain a better perspective we will be better able to place the importance of uncertainty in its proper statistical context, which will aid in better designing approaches that reduce that uncertainty. Thus, in using this approach it is important to do so by initially using these methods as simple guidelines or benchmarks, until this uncertainty is better understood and the rules for addressing the uncertainty to provide the "best approaches" are fully developed.

2) **EPA should consider the following attributes for selection of compounds used in the future development of AOPs:**

   - Wide range of species sensitivity;
   - The most sensitive species is not an MDR species;
   - Varied probit slopes indicating a wide range of sensitivities;
   - Variable species sensitivity slopes;
   - Chronic effect data linked to molecular initiating events (otherwise it may be necessary to collect additional chronic data);
   - Compounds with more than one AOP;
   - AOPs that include salt water responses (e.g., salinity-osmoregulation);
   - Measured data concentrations, not nominal concentrations;
   - Allow for the development of practical guidance for the use of EFs;
   - Importance of developing dose-response relationships that enable estimates of the effects at low doses with acceptable confidence intervals; and,
   - Has adequate chronic effects data (with and without the same acute and chronic endpoints) including important molecular endpoints (e.g. proteomics, metabolomics and Tox21 endpoints).

<u>Additional Comments</u>

Below are some additional comments on materials presented to the Panel.

1) **Statistical considerations**

From a statistical point of view, consideration of MOA can decrease uncertainty of HC5s and especially those estimated using EFs. MOA knowledge may be used to specify which organisms to measure and what endpoints to measure. Moreover, MOA knowledge may help specify which event in the AOP is used as the critical event. These are design not analytical issues that can have a significant effect on estimate uncertainty.

2) The importance and relevance of resident versus non-resident species

The sensitivity of the six AChE-inhibiting chemicals was considered in terms of taxonomic variability (all species vs. resident species only). Only 50% of the resulting combinations were selected for more detailed analysis and out these six AChE-inhibiting chemicals, five used resident species and one used all species (e.g., for diazinon). The rationale for this use of all species for diazinon was that the resident species-only data had a FAV of 0.23 µg /L which was two-fold lower than the GMAV of 0.40 µg/L. Thus, the use of the all species data was advocated for other AChE inhibitors under these types of conditions when the FAV was well below the GMAC (Appendix D). Yet in the white paper, the EPA stated that the "similarity of median EFs for the two MOAs analyzed can be attributed to each MOA having characteristics that lead to a high probability that the lowest toxicity value is near the reference HC5." These two contradictory statements left some of the Panel members confused.

Based on the analysis, aquatic invertebrates were the most sensitive species to AChE inhibitors. Resident invertebrate tests were shown to directly affect the FAV and HC5. The Panel questioned the rationale for adding non-resident, presumably less ecologically relevant and less sensitive species to calculate EFs. The addition of non-resident species may increase the sample size and may enhance the overall power of this statistical approach, but have no ecological relevance. For the narcosis MOA this was not the case, the addition of data from the non-resident species, zebrafish and/or Japanese medaka, was beneficial for chemicals having this MOA, e.g., narcotics.

3) The importance of using the most sensitive species in risk assessment

The most sensitive species may differ from pesticide to pesticide and from one MOA to another. For example, the most sensitive marine species to endosulfan is the pink shrimp which has a LC50 value an order of magnitude lower than other invertebrates tested with endosulfan including other Penaied shrimp species and grass shrimp. Use of pink shrimp as the most sensitive species has resulted in a low (highly protective) saltwater aquatic life criterion value of 8.5 ng/L. This concentration has served as a benchmark in protection for this highly sensitive marine species and other marine life. The Panel cautioned EPA to be certain that the use of EFs strengthens (not weakens) the use of most sensitive species and relevant chronic endpoints as benchmarks for aquatic life protection.

4) Understanding the importance of individual species in each MDR in developing AOPs

The EPA's white paper and Appendix D did a good job in analyzing the effect differently configured MDRs have on estimates of EF and HC5 values. In the white paper (Table 3, p 16), the Panel saw how EFs produce estimated HC5 values that are 2.5-4.5 times lower that the actual FAV for each of the three pesticides examined. This suggests that EFs strengthen results for most sensitive species tests, and in particular for studies using invertebrates. Similarly, in reviewing the MOA data from Appendix D, the Panel noted that using more than four species in MDRs does not appreciably reduce EFs. This is shown for EFs based on the composite geometric mean and those based on the composite 75 and 90^{th} percentiles using AChE inhibitor data from two

fish and two arthropod test species (Appendix D, Figure 2, p 19; Table 4, p 20). Requiring two invertebrate tests in the MDR resulted in less variability and lower EFs. The Panel concluded that the three aquatic species required by OPP are very important, while the additional species included in the MDR do add value. Identifying invertebrates as the most sensitive species and most vulnerable part of the ecosystem helps guide future field monitoring/assessments to confirm that predicted laboratory effects persists in the environment. The MOA approach allows determination of which combinations of species provide the greatest support for an ecological risk assessment. The value of particular species in the MDR used to estimate the HC5 is likely to be MOA specific. Knowing the value that specific species contribute to the assessment of impacts for a specific MOA allows optimization of MDRs for specific classes of chemicals having similar AOP and in particular selection of which aquatic species are tested as part of future pesticide registrations.

5) Understanding the importance of intra-species sensitivity

The EPA's white paper clearly addresses the importance of inter-species sensitivity, but does not address intra-species sensitivity differences. The Panel indicated that intra-species sensitivity differences are an important factor to consider when developing MOA-specific EFs. One panel member provided an example illustrating the importance of intra-species sensitivity differences. Studies by NOAA and EPA showed that LC50 values for grass shrimp tested with endosulfan were much lower in populations collected from the coast of South Carolina than from populations collected from the west coast of Florida (G. Scott, personal communication). Subsequent testing of the grass shrimp populations using microsatellite and other molecular approaches found major genetic taxa differences in the two geographically distinct populations. These results suggest that there might be a genetic basis for the toxicological differences in these two populations. As more molecular based information becomes available for more and more species, these intra-species differences will become better understood and used in their appropriate ecological context in future risk assessments.

(b) Please comment on the strengths and limitations of both the resampling and distributional methodologies for developing MOA-specific EFs, and provide suggestions on their further development and application to other MOAs. Please provide suggestions for deriving EFs for chemicals with MOAs that have limited amounts of empirical toxicity data.

Background

The Host et al. (1995) approach uses random estimates of both the EF and the HC5 by resampling for individual species. The use of random resampling is a sound approach with two estimates – both the EF and HC5. The new distributional approach is a modification of the Host et al. (1995) approach that only estimates the EF by random resampling and uses the entire data set of the AOP for all species to estimate the HC5. Both resampling and distributional approaches produce a "distribution of EF values." The Host et al. (1995) distribution describes the multiplier needed to take the minimum of a subset of a sample of specified size (e.g., N=8) to the HC5 estimated from the full sample. In EPA's analysis, the distributional approach describes

the multiplier needed to take the minimum of a subset of a sample of specified size to the HC5 estimated from the full database of available GMAVs.

General Conclusions and Recommendations

The Panel recommended that the EPA maintain the flexibility of using both approaches to address the variety of MDR data sets for chemicals in a given MOA. The Host et al. (1995) method provides a process using random estimates of both the EF and the HC5 by resampling. The use of random resampling is a sound approach, although it provides a method where there are two estimates – both the EF and HC5. There are compounded, conditional probabilities with the two estimates using this approach that may contribute greater uncertainty to estimates made using this approach. The distributional approach, which modifies the Host et al. (1995) method, only estimates the EF by random resampling, uses the entire data set of the AOP to estimate the HC5, and is not confounded as much by this conditional probability and compounded error question of the original Host et al. (1995) method. Thus, this method should result in less uncertainty and has the advantage of being probabilistic.

The EPA's analysis in Appendix D indicated that the EFs were similar when using these two different approaches (resampling vs. distributional) despite the marked differences in the two toxicity data bases used for computation of results (p 26 in the white paper). For example, the median EFs generated using data from one daphnid and one fish differed by no more than a factor of 1.3; whereas, at the 90^{th} percentile they differed by a factor of 1.4. These are remarkably similar results. The EPA provided additional support for the distributional approach in their presentations and in Appendix D. The Panel concurred with the EPA's findings specific to the AChE AOP; however, they stressed the importance of flexibility in choosing which approach might be best for development of EFs and HC5 because of the variety of MDR data sets for different chemicals with the same MOA.

1) **Resampling approach**

The Panel discussed the strengths and limitations of the resampling approach. The scenarios described in Table 2, Appendix D, are a reasonable set of incomplete MDR-compliant data set conditions which can be used to examine the effectiveness of EFs. The EF distribution computed using the Host et al. (1995) approach uses a specific sample-based HC5 target of the full dataset based HC5 target. The Panel stressed the point that not every scenario needs to be evaluated at once. Limiting the dataset to those chemicals that have at least 15 genera and where the reference HC5 is >80% of the lowest GMAV produces more favorable scenarios where EFs are most likely to be realistic and moderate in value.

The Panel described several limitations of the resampling methodology. Samples can only be constructed of combinations of values in the data base. If the data base inadequately describes the range and distribution of potential GMAVs, then the resulting analysis will be inadequate as well. There may only be a limited number of MDR compliant unique samples possible from a small initial database. The resulting distribution of EF values may not represent the full range or distribution of potential values. If there is one extreme GMAV value in a fairly limited full dataset, then that value would likely have a large effect on the distribution of EF values. The

Panel noted that Appendix D did not provide guidance on what constitutes an inadequate data base.

2) Distributional approach

The Panel provided comments on the strengths and limitations of the distributional approach. The distributional approach starts by developing distributions for the individual components or "classes" of the MDR. Once developed, each distribution essentially defines the range and distribution of potential GMAV values for that class of taxa. Extensive sampling from the individual class distributions in proportion to their assigned representation by the MDRs or pesticide testing guidelines produces a good estimate of HC5 from the "full data set," otherwise referred to as the "composite data set". Similarly, the Panel noted that the Host et al. (1995) or target HC5 estimates in Appendix D for generating the EF distributions can be implemented.

The distributional approach allows creation of a "composite" data set that can be sampled to produce the full range of possible data sets that are MDR compliant. The HC5 estimate derived from extensive sampling of the "composite" distribution is likely to be a good estimate of the true HC5 value if it were possible to sample a large population of species. The "composite" distribution will be smooth all the way out into the tails of the distribution. Unless the sampling specifically assigns extra probability (sampling weight) to extreme values, extreme values are not likely to factor greatly in the estimate of the HC5. The Panel stated that the results of the analysis depend very much on how the individual distributions for each MDR are generated, their extent of smoothing, and the degree to which the final distributions capture the essence of the true distributions, for example, the degree of skewness, kurtosis, multimodality, variance, etc.

The Panel made the following specific comments on the two approaches:

- ***Allow this work to be extended to more MOAs and associated AOPs.*** The Panel unanimously agreed that more chemicals from different MOAs should be analyzed. The distributional approach appears to be more appropriate for a wider range of possible responses than does the resampling approach. This is an important point because the distributional approach has the potential to support EF determinations for more MOAs and associated AOPs. Using data from more chemicals within the MOA with similar AOP will lead to more robust statistical analysis. It is thus not limited to just those data sets that all MDR for ALC and are large enough for proper resampling approaches to be employed. The Panel pointed out that the resampling approach would require more data than the distributional approach. The distributional approach can influence which category to use (Appendix D, Figure 5). For at least two MOAs, the Panel indicated that there is the ability to go through the process and decide which species requirement categories are critical for the analysis. It is important to understand the limitations of the data and how to work within these limitations.

- ***Choose species not at the extreme end.*** The Panel recommended that the EF should not be based on a species with sensitivity much greater than the next most sensitive species. An example where this is an issue is fenitrothion where the most sensitive species is 50-fold more sensitive than the next most sensitive species. Public commenters (representing

CropLife America) suggested the use of median values rather than those from the sensitive tail percentile for determining the EF in those situations where species display high variability in sensitivity. The Panel agreed that the EPA should consider this suggestion. They emphasized the importance of thoroughly examining these data and using best professional judgment in deriving EFs.

- ***Resampling approach should be further tested with other MOAs.*** The resampling approach should be tested for more MOAs to see if similar results are obtained when comparing the narcotizing agent MOA to the AChE inhibition MOA. Narcosis is a more generalized response than AChE inhibition and thus may involve more than one MOA.

- ***Additional statistical considerations***. The current method of computing the EF uses the minimum observed GMAV in the sub-sample as the reference value. The minimum is a statistic with very "poor" properties, extremely skewed sampling distributions, and associated high variability. Instead of the minimum, a percentile estimated from the subsample, such as the HC5, HC10, or HC50, might be used as the *f*(smaller toxicity dataset) in the EF estimation equation. This substitution will produce EF estimates with sampling distributions having much better statistical properties, e.g., smaller variance, which results in less estimate uncertainty for the resulting predicted HC5 value.

  Following the initial discussion of this charge question, the EPA asked the Panel for additional suggestions on how to derive EFs for chemicals with MOAs that have limited amounts of empirical data. One Panel member suggested that within the Bayesian formulation it may be possible to draw information or "borrow" knowledge from other similar MOAs to help settle on an EF value that is not necessarily the default EF. Combining MOAs, as was done for narcosis, may provide more data points that would better refine the EF across multiple groups and taxa, e.g., narcosis. Some MOAs are more specific, making this difficult to do, but other MOAs do not have this limitation.

  The Panel also indicated their general support for the use of a hierarchical approach as described by presenters representing CropLife America. They recommended that additional information to determine trends across multiple MOAs should be considered. The Panel also noted that there were differences in convergence patterns between the two MOAs studied.

- ***Convergence considerations***. The Panel suggested that the EPA consider modifying the models typically used for low dose extrapolation. Research in this area has developed a number of methodologies that effectively use small amounts data to estimate percentiles as low as 5%. Also suggested was the use of modeling techniques (unspecified in the discussion) other than the distributional approach to take advantage of the different amounts and types of data. These alternate modeling techniques may facilitate development of multipliers having different form and properties than the EFs discussed in the white paper. Where data from fewer than 20 species are available, some form of extrapolation might be used.

- ***Importance of reporting confidence limits with EFs, HC5 and other estimates***. Several panel members suggested that computing confidence limits for estimated EF and HC5 values is important. The confidence limits are lower when using the entire data set than when using a subset of values for both the resampling and the distributional approaches. Understanding the range of confidence limits will help set the context for how much value and certainty is placed on these estimates.

- ***Testing of other AChE inhibitors for chronic affects, e.g., fenoxycarb***. The Panel recommended that the EPA consider testing fenoxycarb with other carbamates and possibly organophosphate pesticides using endpoints that are not directed related to AChE inhibition such as reproduction and narcosis. Fenoxycarb disrupts invertebrate reproduction and causes narcosis (Banks1988; Dee 1988; Lee and Scott 1989), but does not inhibit AChE. Study comparisons could be driven by results of (Q)SAR modeling and AOP conceptualizations.

c) Please comment on the strengths and limitations of the different approaches in the Host et al. (1995) and the current work for assigning reference HC5s and summarizing EF distributions.

Background

The Host et al. (1995) method (referred to here as the Host approach) uses random resampling for individual species to estimate both the EF and the HC5. The distributional approach, a modification of the Host approach, uses random resampling to estimate the EF, but uses the entire set of all species data to estimate the HC5. Both the Host and distribution approaches produce a "distribution of EF values." Under the Host approach, the EF sampling distribution describes the multiplier needed to increase the minimum of a subset of a sample of specified size (e.g., N=8) to the HC5 estimated from the corresponding sample from which subsets are drawn. Under the distributional approach, the EF sampling distribution describes the multiplier needed to increase the minimum of a subset of a sample of specified size to the HC5 estimated from the full database of available GMAVs.

General Conclusions and Recommendations

The Panel compared the strengths and limitations of the distributional and Host approaches. The estimated EFs derived using the distributional approach had better statistical properties than the EFs derived using the Host approach. The Host approach derives an EF estimate that has a sampling distribution that is shifted to the right (biased to larger values) and has greater spread (higher variance) than the sampling distribution for EF estimates computed by the distributional approach (see Table 6, Appendix D). The Panel noted that this difference was due to the single HC5 estimate (for this analysis the estimate is a constant value) derived from the complete data sets used in the distribution approach. The Host approach uses a "current effort" mean EF value that is derived from a smaller sample, one that is less likely to include genera with intermediate sensitivities. The Host approach estimates vary from simulated data to simulated data; hence, they are themselves random variables. The variability in the EFs derived using the distribution approach is only related to the sampling distribution of the minimum HC5 estimated from the

small sample subset. The Host approach reports the sampling distribution of a ratio of random variables, which integrates the sampling variability of the HC5 estimated from the small subsample with the sampling variability of the reference HC5 estimated from the selected sample of specified size. Statistically, the Host approach's EF estimate sampling distribution will always display greater variability than the distribution approach's EF estimate sampling distribution.

In Table 6 (Appendix D), the EPA compared the upper 10^{th} percentiles of the EF distribution from both the distributional approach and Host approach. The difference in the EF estimate sampling variability and bias results in the Host approach derived EF was two or more times greater than the distributional approach derived EF for AChE-inhibiting chemicals.

Overall, the Panel agreed that the EPA's analysis supported the conclusions of Host et al. (1995), i.e., reasonable HC5 estimates can be derived from data sets that may be lacking in some species needed to satisfy an aquatic life MDR. In addition, it is possible to quantify the uncertainty of these estimates through use of a resampling simulation.

No method weaknesses were reported.

Specific Comments

The Panel determined that there were sufficient data for the six AChE-inhibiting pesticides to satisfy all aquatic life criterion MDRs for each pesticide. These data were used with random resampling to develop EF sampling distributions for scenarios that met from two to all eight MDRs used to set an aquatic life criterion. Limiting the scenario to requiring at least one planktonic crustacean test and one fish test allows for the estimation of EF values with acceptable characteristics (Figure 1, p 25, white paper).

The Panel noted that using MDRs with more than two fish and two arthropod test species did not appreciably reduce EFs assessed by examination of the composite geometric mean (see Figure 2, p 19, Appendix D) or the composite 75 and 90^{th} percentiles (see Table 4, p 20, Appendix D). The MDRs with two invertebrates resulted in lower mean EFs and less variability in their associated sampling distribution than a MDR that had one. The invertebrate species tested to meet the pesticide data requirements are also the species most sensitive to AChE inhibition. For other MOAs, a broader and likely different set of test species would be needed to constitute an optimal (small) MDR.

The following comments are similar to those made previously in response to Charge Question 4(b).

1) **Importance of reporting confidence limits with the estimates of EFs and HC5 values**

As stated in the Panel's response to Charge Question 4(b), inclusion of confidence limits with estimates of EFs and HC5 values provides context to the certainty or uncertainty of EF and HC5 estimations.

2) Importance of saltwater species

Inclusion of saltwater species might increase the certainty in estimates of EF and HC5 values due to enhanced statistical power of the combined data sets. This was demonstrated for the AChE inhibition MOA. This may not be the case for other MOAs or AOPs, e.g., synthetic pyrethroids.

In examining other MOAs, the Panel thought it might be useful to consider pesticide-salinity interactions in guiding development of future AOPs. If pesticide-salinity interactions do not indicate a statistically significant effect on sensitivity of freshwater and saltwater species, then combining the data would be logical. On the other hand, if there are divergent results, then concerns about the importance of salinity would need to be resolved through additional testing.

The Panel noted that the data presented in Appendix D should allow an assessment of the extent to which the freshwater MDRs and saltwater MDRs were satisfied. The EPA responded to the Panel's query on this matter and indicated that the saltwater requirements were more often met than the freshwater requirements.

3) Most sensitive species considerations

The Panel encouraged EPA to use "sound science" and incorporate the use of these sensitive species across all programs that assess aquatic risk of stressors. The use of more representative sensitive species will lead to HC5 values that are more protective of all aquatic organisms. *C. dubia* and *H. azteca* are discussed as representative "most sensitive species" in Charge Question 4(b). *C. dubia* is recommended for whole effluent toxicity (WET) testing and *H. azteca* for water column and sediment exposure testing. Other species discussed were Unionid mussels (ammonia toxicity) and snails (metal toxicity). A recent paper by Brix et al. (2011) indicated that field concentrations showed effects at lower concentrations than previously measured under laboratory conditions.

4) Consideration of other physicochemical interactions that may affect saltwater and freshwater species differently

There may be other factors such as pH and temperature that may have an impact on an AOP. The Panel recommended that the EPA further expand on the idea of interacting factors affecting toxicity tests under different AOPs. For example, pyrethroids are often more toxic at cold temperatures as they have inverse temperature coefficients. Pyrethroids were more toxic to *H. azteca* at lower temperatures (Coats et al. 1989). Other factors such as pH will affect both physiological responses of freshwater and saltwater organisms and the chemical speciation of some compounds, e.g., effects of these factors on trihalomethanes (THMs) such as bromine (saltwater) (Scott and Vernberg 1979; Scott 1982) vs.chlorine (freshwater) analogs. Hypersaline conditions also enhance the toxicity of aldicarb, azinphosmehtyl (Scott et al., 1987; Fulton and Scott, 1989), fenthion, and phorate to euryhaline species (Lavado et al. 2009; Lavado et al. 2011; Wang et al. 2001). Time to death following chlorpyrifos exposure is also significantly enhanced in salmonids acclimated to hypersaline conditions (Schlenk unpublished). The Panel encouraged EPA to identify this information and use it for future AOP development.

5) Climate Change

The Panel noted that interactions from climate change will be an important consideration. Biological complexities related to climatic changes may result in increased in-stream flows that are nutrient rich. There may be more dramatic seasonal and diel fluctuations in pH, particularly in nutrient-enriched surface waters, which will result in additional uncertainties for ionizable chemicals (Valenti et al. 2011).

Charge Question 5: Use of ACRs to estimate chronic toxicity endpoints or HC5 values for acutely sensitive species

Chronic toxicity data are generally available for fewer species than are tested to determine acute toxicity. Acute-chronic ratios (ACRs) are a common tool used to estimate chronic effect thresholds for taxa lacking chronic toxicity data. Both OPP and OW currently use ACRs developed for specific chemicals to estimate chronic effect levels. ACR distributions have been developed for broad groups of chemicals without consideration of MOA or taxa (e.g., mixed pesticides by TenBrook et al. 2010; all chemicals by Host et al. 1995). ***Appendix E*** *of the white paper considers the application of ACRs developed for chemicals within the same MOA to estimate chronic effects for chemicals with the same MOA and lacking any chemical-specific ACRs. Analysis of ACR data specifically for AChE inhibitors identified patterns in the ACR distributions. Most notably, ACRs for invertebrates were smaller than those for fish. Although fish had larger ACRs, this did not translate to higher chronic sensitivity of fish to AChE inhibitors, because of their very low acute sensitivity. In fact, analyses showed that concentrations protective of acute toxicity to invertebrates would simultaneously protect fish from chronic effects. By extension, for AChE inhibitors, applying an invertebrate-based ACR to the acute HC5 (i.e., fifth percentile of a sensitivity distribution) should provide a chronic HC5 protective of both fish and invertebrates. When comparing ACRs developed for broad groups of chemicals and taxa to ACRs for aquatic invertebrates exposed to AChE inhibitors, they were approximately 2 fold different.*

Panel Response

a) Please comment on the strengths and weaknesses of using chemical-specific ACRs to estimate chronic effect thresholds for other species and taxa.

The Panel was encouraged to see such cross-office collaboration within the EPA to address the timely topic of ACR derivation. These efforts are a great step forward in developing consistent methods for evaluating aquatic risks to stressors across the Agency. The ACR approach illustrated the ability to assess the importance of factors ranging from fish vs. invertebrate species sensitivity differences, freshwater vs. saltwater species sensitivity differences and differences for organophosphates vs. carbamates in developing ACRs. Examples discussed in the presentation and white paper show the overall utility and flexibility of this approach.

The major strength of employing the chemical-specific ACRs approach is that it likely represents the most robust concept at the present time to address chronic toxicity data scarcity issues

(Kenaga 1982; Sloof et al. 1986; Rand 1995; Länge et al. 1998). However, it may be warranted to simply perform additional chronic studies instead of estimating chronic thresholds particularly for mechanisms/MOAs, organisms and chronic responses that are data poor at the present time. Uncertainties for such limited datasets may be too large to overcome without additional chronic adverse outcome data. The AChE inhibitors used in this exercise represented an excellent initial effort due to the amount of data available for this common MOA. Future studies should employ AOPs to identify appropriate chronic endpoints in various organisms prior to ACR derivation, particularly for specifically acting chemicals. For example, there is a need to assess more AOPs associated with pyrethroid insecticides (Ankley et al. 2010).

Developing ACR values that span across pesticides with different MOAs is not recommended, although this had been done previously for industrial chemicals. It remains important to not presume that the MOA causing the acute effects is the same MOA for various chronic effects. Therefore, it would be important to select appropriate chronic endpoints for ACR development. A similar approach was proposed by the EPA Science Advisory Board convened by OW (USEPA 2009). For example, use of *in vitro* screening values using a ligand equivalency approach (e.g., toxicity equivalency) could be used to identify potential AOPs and then select chronic endpoints to strengthen ACRs development (Ankley et al. 2007).

The Panel indicated that the practice of generating a mean ACR value across all available chronic endpoints, which may result from different AOPs within and among species, is not appropriate. For example, coupling fish early life stage and life cycle values were combined for this analysis, albeit to illustrate the relative greater sensitivity of invertebrate acute response thresholds than fish chronic thresholds for AChE inhibitors.

Appendix E did not provide a summary of the various fish species or endpoints used in this analysis, although the presentation did provide a data table to identify the data for several fish species examined. The Panel noted that several studies have recently shown that fish (particularly salmonids) developmental effects, growth impairment and potentially reproduction responses to AChE inhibitors results from a mechanism (i.e., olfaction) that disrupts feeding behavior (Scholz and Hopkins 2006). Thus, these chronic effects may not be related to AChE inhibition at all, but likely occur through signal transduction impairment at lower concentrations than those that inhibit AChE. Consequently, employing these endpoints for ACR values would result much larger estimates. Further, Ahlers et al. (2006) clearly showed that ACRs for all chemicals are higher in fish than invertebrates. Fortunately, additional fish chronic data, particularly for fish life cycle assessment, will be available from endocrine disruption screening and testing efforts. These additional data will augment future analyses of fish chronic response for ACR derivation.

The Panel pointed out that some chemicals have multiple MOAs which may elicit acute and chronic toxicity through different pathways within a species and certainly in other taxa. Here again, selection of species and associated chronic responses should carefully consider AOP when employing ACRs to predict thresholds in other species. Ongoing studies as part of the Tox21 screening program (http://www.epa.gov/ncct/Tox21) with *in vitro* and zebrafish models will assist in identifying other chemical MOAs or unintended/unexpected side-effects that may result in adverse outcomes.

The Panel also noted that the toxicity data used were from studies in which the exposure concentrations were both nominal and analytically verified. Good science would dictate use of analytically verified concentrations whenever possible, though some historical studies contain nominal data and may represent the only such data for some compounds. Similar comments apply for using NOEC and LOEC values from hypothesis testing to develop ACRs. The Panel was very encouraged to see the EC20 analyses included in Appendix E.

Comparison of the different approaches indicated different values of each distributional endpoint considered at the 50^{th}, 80^{th}, and 90^{th} centiles, although the differences were less than a factor of two at each of the three percentile estimates assessed. These results suggest that there is consistency in the overall approach. EPA and public commenters (from CropLife America) pointed out that choice of distribution model affects the HC5 estimates, which in turn affects the ACR derived from these estimated. The method for calculating the ACR presented in Appendix E shows that propagation of error could be a significant source of uncertainty in the resulting ACR estimate. The ACR appears to use two sets of extrapolations, one for HC5 and the other for the ACR itself. The form of these extrapolation equations dictates that input values should not be truncated prior to input (i.e., not reduced to a pre-specified number of significant digits). Another comparison would involve examining ratios of confidence limits associated with the HC5 estimate and with the ACR estimate. The Panel recommended that resampling approaches, e.g., two-dimensional Monte Carlo simulation, be considered.

b) Please comment on the strengths and weakness of applying "default" ACRs derived from other chemicals to extrapolate from an acute HC5 to a chronic HC5, including the relative merits of values derived for MOA-specific (e.g., for AChE inhibitors) or more generalized (e.g., Host et al. 1995) distributions.

As noted above in response to Charge Question 5(a), the Panel recognized that applying "default" ACR values is necessary to account for uncertainties based on the paucity of available chronic toxicity data for most industrial chemicals. The Panel agreed that application of sound scientific principles to derive ACR uncertainty factors that may be applied to new substances or data poor chemicals was a reasonable approach. When possible, however, MOA-specific ACR values, which are influenced by various chronic AOPs, should be developed by endpoint in major taxa (e.g., macrophytes, algae, invertebrates, vertebrates) using AOPs to avoid under- or over- estimation of chronic thresholds. The Panel re-emphasized a point made in response to Charge Question 5(a) that employing default ACR values and those derived by averaging ACR values across chronic endpoints within and among species is not recommended, particularly for biologically active molecules because various chronic responses likely result from different AOPs. At best, in the absence of these considerations, the Panel recommended the development and application of uncertainty factors to ACR values calculated across endpoints and species for biologically active chemicals.

The Panel was supportive of the generic approach proposed in the presentation made by a public commenter (representing CropLife America) to develop a hierarchy of preferred data attributes during ACR development (slide 31, CropLife America presentation, see EPA public docket EPA-HQ-OPP-2011-0898).

Chronic endpoints are generally related to growth and reproduction; therefore, care should be taken when extrapolating acute to chronic ratios across species that have different life history strategies. Chronic responses to contaminants have different effects on species that employ different life history strategies (Spromberg and Birge, 2005). Therefore, the Panel recommended that to use an ACR that it be used only for species that employ similar life history strategies (e.g., short life span, fast growth, high reproductive effort vs. long life span, slow growth and reproductive effort spread over several years).

The Panel out that there is a need to examine taxonomic differences during AOP development to account for differences in sensitivity of freshwater and saltwater fish to a chemical and invertebrate sensitivity to the same chemical. The Panel pointed out that pesticides entering marine waters through runoff often have pesticide-salinity interactions which may enhance pesticide toxicity. Thus, while it may be appropriate to evaluate toxicity distributions for acute and chronic comparisons, pesticide - salinity interactions that influence toxicity threshold may occur, resulting in uncertainties, which the white paper did not address.

Observations for AChE inhibitors in this report represent important ecotoxicological findings to explore and allow for creation of larger data sets that are more robust and have greater statistical power. It is important to generate a set of rules or operational guidelines which are provisionally used in future AOP development. For example, comparisons of taxonomic differences and similarities are important in guiding decisions on development of rules. For example, rules for selection of chronic endpoints within an AOP are critical for the development of robust ACRs. These rules should be adaptive and subject to modification as more information and insight is gained. It is clear, however, that employing mechanistic endpoints can reduce uncertainty during ecological risk assessment of biologically active compounds such as pesticides and pharmaceuticals (Brain and Brooks 2012).

c) *Are there other methods for estimating a chronic HC5 that the Panel believes would be technically superior to ACR-based approaches? For example, TCE models (described in USEPA 2010a), distributional approaches (e.g., de Zwart 2002, Douboudin et al. 2004; described in USEPA 2010b).*

The Panel commented on the use of distribution approaches, TCE models, and other approaches for estimating a chronic HC5 other than an ACR-based approach. The Panel indicated that distributional approaches showed great promise, particularly for developing ACR values for specific chemicals with differences between HC5 values derived from acute and chronic species sensitivity distributions (de Zwart 2002). However, data availability will inherently limit the broad implementation of this approach. Further, due to potential differences in chronic AOPs among species, ACR values can vary greatly among species and chronic endpoints selected.

Another possible approach would be to explore the utility of deriving ACR values from differences between centiles from species and chronic endpoint specific chemical toxicity distributions for chemicals with a common MOA, as was recently demonstrated for AChE inhibitors (Berninger 2011; Williams et al. 2011). But here again, data availability limits broad application of these techniques at this time. Although time-concentration-effect (TCE) models

may be useful for some chemicals, the robustness of these models is limited when chronic MOAs differ from acute MOAs, resulting in a chronic response not predicted by TCE relationships.

Use of newer and more refined MOA assessments are needed employing "omics" and "in vitro" testing strategies, such as those being used by the National Toxicology Program and its Federal partners for informing human health risk assessments. These qualitative studies can drive the in vivo studies of specific AOPs needed for assessment of chemicals. For example, it may not be necessary to evaluate the acute toxicity of a particular chemical in eight species if the most sensitive toxicity endpoint (i.e., the endpoint of concern) is a chronic toxicity endpoint. The Panel was unaware of any other approaches to consider in response to this charge question.

The Panel provided the following specific editorial comments on Appendix E:

- Page 4: Is the AquaChronTox database available?
- Page 4: A reference should be inserted to support the statement, "In general, one expects organism weight to vary with the cube of organism length (assuming similar body morphology), so even very small changes in length would be expected to cause substantial changes in weight."
- Page 5: Additional definition should be provided for the logistic regression modeling (e.g., were 3-parameter logistic models used?), as employed using TRAP Ver 1.02: "A sigmoid model with finite tails was used to fit the exposure response data for individual endpoints, except in a very small number of cases where this model did not seem to reflect the shape of the underlying data, in which case a piecewise linear model was used."
- Page 9: "EF" should be defined at its first usage.
- Page 10: "it seems reasonable to base an ACRHC5 on an ACR selected from the joint distribution of ACRs for all invertebrate species." The Panel concurs given that other freshwater invertebrates could be more sensitive than cladocerans to AChE inhibitors. However, fish responses, particularly chronic reproduction thresholds, should be given further consideration.
- Page 11: "As such, it is highly uncertain whether the same conclusion would be reached for another mode of action with different taxonomic sensitivity." In fact, it is highly unlikely that the same conclusion would be reached for specifically acting chemicals eliciting toxicity through other MOAs (e.g., receptor mediated events) as this would depend on the presence of the toxicological target in other nontarget organisms.
- Page 11: "One important consideration is that because ACRs are ratios and both the numerator and denominator have uncertainty, variability in the ACRs from chance alone will be larger than the variability in either of the component values." Quite possibly so, but this will depend on data available for each chemical.
- Page 11: Additional methodological details should be provided for the Monte Carlo simulation.
- Page 14: Are all references cited in Appendix E readily available online? If so, please update with website location information.
- Figure 2. Would not it be useful to compare ACR approaches to plot those tabulated from NOECs, GM and EC0s?

- Figure 6: Fish ELS and life cycle data are plotted on the same figure. This is problematic due to different endpoints, different exposure durations and different AOPs involved in these chronic responses.
- Figure 7 legend: ACR_{GM} is mislabeled in the figure legend as ACR_{ChV}.
- Figure 9 x axis: GM is mislabeled ChV.
- Table 5. Why use NOEC and LOEC? A previous SAB report and recent literature have been critical of this approach.
- Table 5. Rationale for using an EC20 should be provided (e.g., instead of another EC_x value).

Charge Question 6: ***Estimating the HC5 for Aquatic Plants***

As with animals, there has been uncertainty regarding the extent to which available data reflect the range of sensitivities that may be present for aquatic plants. This effort focuses on the potential uses of EFs and SSDs to estimate the HC5 relevant to aquatic plants. As described in ***Appendix F****, the primary analyses investigated the use of the standard aquatic plant species submitted to fulfill FIFRA data requirements in estimating the HC5. These species, which are referred to in the white paper as "the FIFRA-5" include: Pseudokirchneriella subcapitata (a freshwater green alga), Anabaena flos-aquae (a freshwater cyanobacterium), Navicula pelliculosa (a freshwater diatom), Skeletonema costatum (a saltwater diatom) and Lemna gibba (the aquatic vascular plant, duckweed). The analysis described in* ***Appendix F*** *suggested that estimating the HC5 using the FIFRA-5 will result in a reasonable approximation of the HC5 of an aquatic plant community.*

Panel Response:

Summary and general comments

In response to this charge question, the Panel offered the following recommendations to further advance the Agency's efforts:

- Develop MDRs for freshwater and saltwater plants;
- Verify the uncertainty (i.e., over- or under-protection) inherent in the use of certain extrapolation factors for a given number of species in a SSD;
- Develop and implement a strategy for systematically replacing data that were collected using suspect methods or analytical procedures, using nominal values or are not based on curve-fitting endpoint estimates (e.g., NOECs). However, the Panel recognized that in cases in which historical toxicity test data may provide the only relevant toxicity test results for specific chemicals of concern, then EPA should use their best professional judgment regarding the inclusion of such data in the risk analysis.
- Characterize the influence of nutrients on plant toxicity test results both in the standard laboratory bioassays used to generate data for regulatory purposes (issues of non-uniformity among bioassays and investigators) and in the field (issues with nutrient enrichment in many freshwater and saltwater systems).
- Consider the addition of aquatic macrophyte toxicity tests with traditional (e.g., shoot growth) and non-traditional but potentially more sensitive (e.g., root growth) endpoints.

Addition of protocols and endpoints should account for differences in the MOA and mechanism of action of specific herbicides and fungicides.

- Consider the addition of a marine macrophyte toxicity test protocol, and potentially other species (e.g., to develop a FIFRA-5 corollary for marine organisms) in the MDR development.
- Determine ecologically relevant endpoints (i.e., those that determine population dynamics, such as growth and reproduction in animals) for aquatic plants and couple them with animal (herbivore) endpoints to understand potential secondary effects (e.g. loss of primary producers resulting in reduced food for herbivores).
- Initiate a systematic research effort to determine the appropriate effects concentration (HCx) to protect plant community structure and function.
- Incorporate knowledge of AOPs in test species selection and data interpretation to reduce the large uncertainties that exist in this process.

a) *Appendix F of the white paper includes an analysis of the use of EFs and SSDs combined with the FIFRA-5 to estimate the HC5 for aquatic plants. Please comment on the strengths and weaknesses of the methods, results and conclusions associated with this analysis.*

The Panel appreciated the tremendous amount of work that went into these analyses and noted that these efforts were a solid foundation for aggressively moving towards achieving the goal of harmonizing efforts, requirements and conclusions within OW and OPP.

The Panel agreed that employing standardized aquatic plant models to develop SSDs and estimate HC5 values was a reasonable approach. However, the lack of aquatic plant data is problematic and makes it difficult to create SSDs that represent the range of toxicity values for aquatic plant species. The ability to accurately estimate HC5 values would be compromised by this lack of data; therefore, setting aquatic life criteria for aquatic plants based on limited data would be highly uncertain. Having minimum data requirement guidelines for aquatic plants would lead to expansion of the available toxicity data for more plant species. The addition of saltwater species would also be important. More data would increase the robustness of the SSDs and lead to more certain estimates of HC5 values that are used to calculate aquatic life criteria. More certain HC5 estimates would lessen the need for the use of EFs which may be overstated.

Data availability from several common species (e.g., FIFRA-5) may limit the robustness of this approach compared to SSDs for aquatic animals. A comparison of the FIFRA-5 with the full data set (Note: a full data set was defined by the EPA as the FIFRA-5 plus 5 other plant species) that the lowest FIFRA-5 effects value was almost always greater than the median HC5 from the full data set. An EF of two was therefore required to ensure plant population protection using the FIFRA-5 data set. The strength of using an EF in this situation is that it minimizes the likelihood of "under protection". On the other hand, a weakness of using an EF is that it might increase the likelihood of "over protection." For example, some of the EF-adjusted HC5s presented in Appendix F of the white paper were less than the "true" HC5 by a factor of up to 8.1. The EPA presenter pointed out that EFs are unnecessary if toxicity data are available for 10 species. The necessity of using an EF is not clear for $n \geq 5$ but ≤ 10 species; an EF of two seems reasonable.

The comparison of the FIFRA-5 and full data sets demonstrated the strength of using the SSD approach with the 25th percentile value of the HC5 (from FIFRA-5 data) is that it minimizes both under- and over-protection. The precision of the HC5 estimate increases as the number of species with EC50s increases and the effect is most evident for situations where 5 or more species are used. One weakness of the SSD approach is evident in the situation where the fitted regression has a non-significant slope (e.g., the slope of the fitted line is essentially 0). Another weakness was illustrated in the EPA's analysis of the FIFRA-5 data for the herbicide, metribuzin. In these data, one of the toxicity values is much higher than the other four and is 15 fold less sensitive to metribuzin than the next highest value. An extremely less toxic "outlier" would result in a higher HC5 estimate, one in which would be less protective of aquatic life. A better understanding of the "outlier" data might offer a solution to this problem. The effect of a single outlier is always reduced with more data. The Panel indicated that alternate estimates of HC5 based on EFs would be needed for data sets with outliers of this type.

Nutrient concentrations and stoichiometries often vary among growth media formulations. Subsequently, nutrient availability and limitation during algae and macrophyte toxicity tests will influence control growth rates, which can influence point estimates of toxicity thresholds and thus introduce uncertainty in ecological risk assessments (Fulton et al. 2009, 2010). Further, nutrient concentrations and stoichiometries in standardized growth media are often not reflective of environmental concentrations or stoichiometries of nitrogen (N) and phosphorus (P), which also introduces uncertainty in extrapolation of such laboratory observations to aquatic ecosystems (Fulton et al. 2009, 2010). Table 1 provides several examples of the influence of nutrients on growth rates for Lemnaceae. Future studies, particularly for non FIFRA 5 data, should account for data quality considerations and work towards developing MDRs.

Issues of nutrient status are very important in saltwater as 67% of the nation’s estuaries are moderately to severely affected by nutrient enrichment. Further, use of an additional saltwater toxicity test species (e.g., a marine alga) to the traditional suite of FIFRA-5 data should help address assessments of nutrient interactions with plant toxicity. Past experiments conducted by NOAA using a mesocosm demonstrated significant interaction of photosystem II inhibition on nutrient enrichment (Delorenzo et al. 1999).

b) Please comment on other aquatic plant species or test endpoints that may be considered in order to generate more certain estimates of the HC5 for aquatic plants.

When considering other plant species or endpoints or normalizing various endpoints to a standard metric, the Panel commented that evolutionary conservation of toxicological targets among species could be used *a priori* to guide selection of species and endpoints for toxicity studies. This approach would reduce uncertainty during ecological risk assessments of biologically active compounds (Brain et al. 2008a, b; Brain and Brooks 2012).

The addition of more and new species may not always be useful. The Panel suggested that selective addition based on knowledge of AOP and plant community interactions may result in choices that improve the SSD. Additional species need to be in the lower end of the sensitivity spectrum. A weighted regression might be useful to focus on that part of the distribution.

Table 1. Some examples of the influences of variations of Hutner's media on Lemnaceae control growth rates.

Lemnaceae	Control Growth Rate (d^{-1})	Media	Reference
Lemna gibba	0.31	Hutner's media	[1]
L. gibba	0.2	Hutner's media with 13.05 mg L^{-1} P	[2]
Wolffia borealis	0.62	33% Hutner's media	[3]
L. minor	0.45	33% Hutner's media	[3]
Spirodela polyrhiza	0.08	33% Hutner's media	[3]
L. minor	0.26	50% Hutner's media	[4]
W. brasiliensis	0.18	50% Hutner's media	[4]
L. gibba	0.229	Hutner's media + 0.14 mg L^{-1} N	[5]
	0.393	Hutner's media + 1.4 mg L^{-1} N	[5]
	0.402	Hutner's media + 14 mg L^{-1} N	[5]

[1] Mkandawire and Dudel 2005; [2] Mkandawire et al. 2005; [3] Lemon et al. 2001; [4] Lahive et al. 2011; [5] [Fulton et al. 2010.

The EPA suggested that additional data with under-represented species may help determine whether existing data sets provide a relevant (protective) distribution for all plant species. The Panel indicated that the addition of a rooted freshwater macrophyte, *Myriophyllum* sp., to the data set would add robustness to the analysis. Further, as noted by EPA, root growth can be a more sensitive endpoint than shoot growth. Based on the limited available data, *Myriophyllum* sp. appears to be sensitive to herbicides. For example, studies have shown that *M. aquaticum* is more sensitive to atrazine than *Lemna minor* (Teodorović et al. 2012). The reason for this greater sensitivity may be elucidated as AOPs for these chemicals are better delineated. The Panel noted

that an OECD guideline for *M. aquaticum* is under development according to the comments made on behalf of CropLife America.

The Panel recommended eastern Pacific brown algae, *Macrocystis pyrifera,* and the estuarine green alga, *Ulva lactuca,* as representative marine macrophytes. A standard bioassay protocol has been developed for *M. pyrifera*. This is an early life stage toxicity test that is routinely used in effluent monitoring in California as required by NPDES permits. There is an EPA (1995) protocol for this species. The protocol assesses spore germination and germ-tube elongation ("growth" not driven by photosynthesis) over a 48h exposure. Spore germination and growth have been shown to be particularly sensitive indicators of water toxicity using this species, but extensive testing with herbicides and fungicides has not been conducted. A similar assay has been developed using *U. lactuca* (Hooten and Carr 1998).

Plant toxicity testing to look at effects on productivity, e.g., biomass, should be included as part of the aquatic plant toxicity testing regimen. Productivity is an important measure of the ability of lower trophic level organisms such as plants to support higher trophic level organisms via the food chain. The Panel considered that coupling plant test endpoints and acute/chronic aquatic animal effects is important, especially for those animals that rely directly on plant productivity (e.g., herbivores). Previous research has demonstrated the importance of this connection when researchers were able to couple new saltwater algal endpoints with chronic bioeffects in mollusks (Dee 1988).

The Panel noted that a protocol assessing growth and chlorophyll *a* production developed using rice (*Oryza sativa*) could be considered to provide additional sensitivity data for vascular plants (Powell et al. 1996). A seed germination and root elongation test has also been developed for cattail (*Typha latifolia*), and this could be considered to provide an additional endpoint using an emergent vascular plant (Moore et al. 1999). However, determination of appropriate aquatic plant species and toxicity endpoints requires knowledge of the chemical's MOA.

c) Please comment on the strengths and limitations of normalizing plant toxicity endpoints to a standard metric (i.e., EC50 for growth rate).

One of the strengths of normalizing plant toxicity endpoints to a standard growth metric is that it allows for comparisons of the relative sensitivity of toxicity test protocols to chemicals and of the relative toxicity of herbicides/fungicides. This was illustrated in the EPA's assessment of the effects of atrazine on a whole plant community (Appendix F).

One weakness of this approach is that measures of growth vary with different protocols. For example, the four FIFRA micro-algal tests measure population growth rate, while assays using vascular plants typically measure growth of individual plants. Other measures of "growth" include root shoot growth, e.g., cattails (*Typha sp.*) or spore germination tube elongation, e.g., giant kelp (*Macrocystis pyrifera*).

The Panel did not recommend the normalization approach for characterizing risk of chemicals to whole plant communities since growth measures alone may not represent impacts on all members of the community. As the Panel noted, growth is not always the most sensitive

endpoint, e.g., atrazine and diuron full data set SSDs for species other than the FIFRA-5, seaweed sexual reproduction and spore germination.

The Panel stated that it would be inappropriate to normalize all endpoints to a standard metric. This is an important issue and one that retards the advancement of this program. The Panel recommended that precise effects concentration (EC*x*) values should be used to be more protective of the specific plant populations and aquatic plant communities at risk. For example, an EC20 or EC10 may be more scientifically supported levels to protect an aquatic plant community rather than an EC50, depending on the endpoint. Based on the lack of available data to analyze or draw conclusions, the Panel recommended a systematic research effort on the appropriate effects concentration to protect plant community structure and function.

d) Please comment on the extent to which consideration of MOA may decrease uncertainty associated with HC5 values for aquatic plants.

The Panel agreed that consideration of the specific MOA of herbicides and fungicides may decrease uncertainty associated with HC5 values such that appropriately sensitive species and endpoints would be considered. For example, the Panel recommended inclusion of C4 plants (e.g., *Salicornia*) for C4-plant selective herbicides and C3 plants (e.g., *Lemna*) for C3-selective herbicides (e.g., nitriles, benzothiadiazole, phenyl-pyridazines). A better understanding of AOPs (Ankley et al. 2010) would support future studies of aquatic plant toxicology (Brain and Brooks 2012).

Panel Final Comments

The Panel appreciated the tremendous amount of work that went into these analyses and commended OPP, OW, and the Office of Research and Development for working together to advance timely and important questions related to environmental protection of aquatic life uses. Continuing this collaborative work promises to lead to development of robust scientific approaches, which can support increased regulatory efficiency and reduce uncertainty associated with environmental management decision making. This work has laid the foundation for aggressively moving towards achieving the goal of harmonizing efforts, requirements and conclusions within OW and OPP.

As repeatedly stated during the public meeting, the Panel strongly encouraged EPA to publish the work presented in peer reviewed journals and to present their work in national and international scientific meetings. The value of peer review and publication in the presentation, release, interpretation, and use of scientific data has been clearly established in the disciplinary fields, and in the courts (e.g., Daubert v. Merrill Dow 1993).

The Panel provided additional comments following the discussion of all of the Agency's charge questions.

1) Numerical benchmarks

The Panel offered several comments on the EPA's efforts to develop numerical benchmarks that could be used to assess the potential for adverse effects of chemicals for which there are not sufficient data to develop criteria per the 1985 Guidelines.

- These benchmarks should be conservative such that contaminant concentrations below the benchmarks are unlikely to result in chemical toxicity, while concentrations above may result in chemical toxicity (i.e., "potential toxicity," see p 3 of the white paper).
- The level of conservatism should not be excessive such that the benchmarks have little or no value in terms of discrimination between these two possibilities.
- Recent studies have suggested that sensitive populations are not being protected by water quality criteria developed through laboratory exposures or traditional approaches (e.g., revised EPA ammonia criteria (Brix et al. 2011); pink shrimp (G. Scott personal communication). Such information indicates that some validation of whether the aquatic life criteria are protective of sensitive populations should occur.
- The benchmarks that are ultimately developed can be used in problem formulation and possibly also in screening-level ecological risk assessment (SLERA), but not in isolation for detailed-level ecological risk assessment (DLERA) or for definitive decision-making. They can be used to develop hazard quotients (HQs), but not risk quotients (RQ). Hazard is the possibility of an adverse effect. Risk is the probability of such an effect. In an SLERA, the benchmarks will have less weight in a weight of evidence (WOE) determination than appropriate toxicity tests, which will have less weight than properly conducted resident community / population evaluations. As noted in the Agency's white paper (pp 3 and 9), "they are not designed to represent community level benchmarks".
- The benchmarks should be developed based on bioindicators (i.e., endpoints for individual organisms that relate to community- or population-level effects such as survival, growth, reproduction, avoidance [the latter for mobile organisms]). They should not be developed based on biomarkers that only reflect exposure or that measure effects at the cellular or subcellular level that cannot be translated to whole organism effects and thus cannot be translated to community- or population-level effects. For example, EPA's aquatic life criteria clearly define limits on chemical exposures which are considered sufficient to preclude unacceptable effects on aquatic communities, not to individual organisms (with the exception of endangered species) or to sub-organism level components of those organisms.
- Regarding the scientifically-defensible use of these benchmarks, one Panel member quoted Aristotle, *"It is the mark of an instructed mind to rest satisfied with the degree of precision which the nature of the subject permits and not seek exactness where only an approximation of the truth is possible."*

2) Use of most sensitive species, *D. dubia* and *H. azteca*

The Panel expressed concern for the variation in selection of most sensitive species for aquatic risk assessment by OPP and OW. The Panel stressed the importance of using the most sensitive species to protect aquatic organisms in its response to charge questions 4 and 5. As stated in these responses, *C. dubia* replaced *D.magna* as the sensitive species of choice for National

Pollutant Discharge Elimination System (NPDES) permitting and WET testing requirements used by OW. The *H. azteca* sediment toxicity assay provides protection for benthic and epibenthic organisms. These organisms have been observed to be particularly sensitive to some pesticides. Nevertheless, neither of the above species is routinely used by OPP in its ecological risk assessment of pesticides to aquatic organisms. The Panel suggested that OPP consider the use of these species.

3) Application of AOPs

The Panel strongly supported the development and application of AOPs during ecological risk assessments of chemicals. When developing the two anchoring observations within AOPs, it will remain important to identify specific sublethal measures associated with linkages between molecular initiation events (anchor 1) and adverse outcomes at the individual and population levels of biological organization (anchor 2). Accounting for evolutionary conservation of targets across species can support *a priori* selection of potentially sensitive taxa and chronic endpoints for advanced study and ecological risk assessment. Specifically, and as noted above, care should be taken to rigorously test the linkages among sublethal measures along toxicity pathways to ensure that they meet the above-stated goal of protecting aquatic communities.

When uncertainties exist for poorly understood AOPs, particularly those associated with relatively understudied and/or underrepresented organisms (e.g., amphibians, bivalves, amphipods, plants) and endpoints (e.g., ecologically important behavior) in existing standardized methods, the Panel recommended that new data s be collected and robust methods developed in targeted areas. For example, research related to ecotoxicogenomics being performed at the Office of Research and Development's laboratory in Duluth (Minnesota) and computational in vitro and *in vivo* (with zebrafish) toxicology studies associated with the Tox21 effort will support developing an understanding of chemical properties and various biological activities (expected and unexpected) associated with molecular initiation events that lead to adverse outcomes related to population-relevant endpoints such as growth, reproduction and survival in different organisms.

Molecular biomarkers can have a role in determining AOPs as has recently been elegantly demonstrated for vitellogenin and for plasma hormone levels in female fish (Miller et al. 2007; Ankley et al. 2008). Using the fathead minnow as a model, a direct linkage was demonstrated from decreased levels of plasma vitellogenin in female fish, to alterations in ovarian tissue morphology (ovaries where developing eggs were visibly impaired by the lack of vitellogenin), to decreased egg production by females. These measured endpoints were modeled using a Leslie Matrix to predict population declines. As additional AOPs that result in population declines are elucidated, they should be incorporated into contaminant models that are used to set benchmarks.

4) Revisions to documents, guidelines, and benchmarks are necessary over time

Testing requirements, guidelines, and benchmarks should not remain static, but should be subject to periodic scientific peer review as stated in the “Good Science Clause” of EPA's 1985 Guidelines (EPA 1985):

> *On the basis of all available pertinent laboratory and field information, determine if the criterion is consistent with sound scientific evidence. If it is not, another criterion, either higher or lower, should be derived using appropriate modifications of these Guidelines.*

A prime example of guidelines and data requirements which should be updated is the 1985 Guidelines (and MDRs) which are now more than 25 years old. There have been substantial advances in the science and new data have been gathered during the intervening years. As part of the ongoing development of common effects methodology, the Panel recommended that a systematic review be conducted on the state of the science in pesticide testing requirements under FIFRA and aquatic life criteria data guidelines under CWA, as well as a critical examination of the quality and validity of the existing data to substantiate scientific conclusions about potential effects of chemicals (including pesticides) to aquatic communities.

REFERENCES

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrano JA, Tietge JF, Villeneuve DL (2010) Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29: 730-741.

Ankley GT, Brooks BW, Huggett DB, Sumpter JP (2007) Repeating history: pharmaceuticals in the environment. Environ Sci Technol 41: 8211-8217.

Ankley GT, Miller DH, Jensen KM, Villeneuve DL, Martinović D (2008) Relationship of plasma sex steroid concentrations in female fathead minnows to reproductive success and population status. Aquat Toxicol 88: 69-74.

Bailer AJ, Oris, JT (1999) What is an NOEC? Non-monotonic concentration-response patterns want to know. Learned Discourse, SETAC News, March, 1999. SETAC Press, Pensacola, FL. pp. 22-24.

Banks JB (1988) The acute toxicity of diflubenzuron, methoprene, temephos, fenoxycarb, *Baccillus thuringiensis* var. *israelienis* (BTI), and a fenoxycarb/BTI mixture on the nontarget, saltmarsh fiddler crab, *Uca pugilator*, with an emphasis on the sublethal effects of fenoxycarb and the fenoxycarb/BTI mixture on energy metabolism. 72 pp.

Bermudez-Saldana JM, Cronin MTD (2006) Quantitative structure-activity relationships for the toxicity of organophosphorus and carbamate pesticides to the Rainbow trout *Onchorhyncus mykiss*. Pest Man Sci 62: 819-831. [DI 10.1002/ps.1233]

Bermudez-Saldana JM, Escuder-Gilabert L, Medina-Hernandez MJ, Villanueva-Camanas RM, Sagrado S (2005) Chromatographic evaluation of the toxicity in fish of pesticides. J Chromatog B – Anal Techol Biomed Life Sci 814: 115-125.

Berninger JP (2011) Leveraging available data for contaminants of emerging concern to develop an understanding of environmental hazard. Ph.D. Dissertation, Baylor University, Waco TX.

Biesiada J, Porollo A, Velayutham P, Kouril M, Meller J (2011) Survey of public domain software for docking simulations and virtual screening. Hum Genomics. 5(5): 497-505. Review. PubMed PMID: 21807604.

Brain RA, Brooks BW (2012) Considerations and criteria for the incorporation of mechanistic sub-lethal endpoints into environmental risk assessment for biologically active compounds. In: Brooks BW, Huggett DB (Eds). Human Pharmaceuticals in the Environment: Current and Future Perspectives. Springer, New York.

Brain RA, Hanson M, Solomon KR, Brooks BW (2008a) Aquatic plants exposed to pharmaceuticals: Effects and risks. Rev Environ Contam Toxicol 192: 67–115.

Brain RA, Ramirez AJ, Fulton BA, Chambliss CK, Brooks BW (2008b) Herbicidal effects of sulfamethoxazole in *Lemna gibba*: using *p*-aminobenzoic acid as a biomarker of effect. Environ Sci Technol 42: 8965-8970.

Brix KV, DeForest DK, Adams WJ (2011) The sensitivity of aquatic insects to divalent metals: A comparative analysis of laboratory and field data. Sci Total Environ 409: 4187-4197.

Casalegno M, Sello G, Benfenati, E (2006) Top-priority fragment QSAR approach in predicting pesticide aquatic toxicity. Chem Res Toxicol 19: 1533-1539. [DI 10.1021/tx0601814]

Celander MC, Goldstone JV, Denslow ND, Iguchi T, Kille P, Meyerhoff RD, Smith BA, Hutchinson TH, Wheeler JR (2011) Species extrapolation for the 21st century. Environ Toxicol Chem. 30(1):52-63. PubMed PMID: 20963850.

Coats JR, Symonik DM, Bradbury SP, Dyer SD, Timson LK, Atchison GJ. (1989) Toxicology of synthetic pyrethroids in aquatic organisms: an overview. Environ Toxicol Chem 8: 671-679.

Daubert et al. v. Merrill Dow Pharmaceuticals, Inc. (92-102), 509 U.S. 579 (1993). U.S. Supreme Court No. 92-102, argued March 30, 1993, decided June 28, 1993.

de Haas EM, Eikelboom T, Bouwman, T (2011) Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR. SAR QSAR Environ Res 22: 545-559.

de Zwart D (2002) Observed regularities in species sensitivity distributions for aquatic species. in: species sensitivity distributions in ecotoxicology. Posthuma L, Suter III GW, Traas TP (Eds). Lewis Publishers, Boca Raton FL.

Dee J (1988) The acute toxicity of four commonly used mosquito control larvicides, a microbial larvicide (BTI) and pesticide mixture on the grass shrimp, *Palaemonetes pugio.* Master Thesis, University of South Carolina, School of Public Health: 63 pp.

Delorenzo ME, et al. (1999) Atrazine effects on the microbial food web in tidal creek mesocosms. Aquatic Toxicol 48:241-251.

Devillers J (2004) Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling. SAR QSAR Environ Res 15: 501-510. [DI 10.1080/10629360412331297443]

Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O (2005) A stepwise approach for defining the applicability domain of SAR and QSAR models. J Chem Inf Comp Sci 45: 839-849. [DI 10.1021/ci0500381]

Duboudin C, Ciffroy P, Magaud H (2004) Acute-to-chronic species sensitivity distribution extrapolation. Environ Toxicol Chem 23: 1774-1785.

El Yazal J, Rao SN, Mehl A, Slikker W Jr. (2001) Prediction of organophosphorus acetylcholinesterase inhibition using three-dimensional quantitative structure-activity relationship (3D-QSAR) methods. Toxicol Sci. 63: 223-232.

Fulton BA, Brain RA, Usenko S, Back JA, Brooks BW (2010) Exploring *Lemna gibba* thresholds to nutrient and chemical stressors: differential effects of triclosan on internal stoichiometry and nitrate uptake across a N:P gradient. Environ Toxicol Chem 29: 2363–2370.

Fulton BA, Brain RA, Usenko S, Back JA, King RS, Brooks BW (2009) Influence of N and P concentrations and ratios on *Lemna gibba* growth responses to triclosan in laboratory and stream mesocosm experiments. Environ Toxicol Chem 28: 2610-2621.

Fulton MH, Scott GI (1991) The Effects of certain intrinsic and extrinsic variables on the acute toxicity of selected organophosphorus insecticides to the mummichog, *Fundulus heteroclitus*. Journal of Env. Sci and Health B26:459-478.

García I, Fall Y, Gómez G (2010) QSAR, docking, and CoMFA studies of GSK3 inhibitors. Curr Pharm Des. 16(24):2666-2675. Review. PubMed PMID: 20642432.

Gupta S, Mohan CG (2011) 3D-pharmacophore model based virtual screening to identify dual-binding site and selective acetylcholinesterase inhibitors. Med Chem Res 20: 1422-1430. [DI 10.1007/s00044-010-9373-7]

Host GE, Regal RR, Stephan CE (1995) Analyses of acute and chronic data for aquatic life – Draft Report. Office of Research and Development, U.S. Environmental Protection Agency, Washington DC. 82 p.

Jaeger T (2012) Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. Environ Toxicol Chem 31: 228–229.

Kaiser KLE, Dearden JC, Klein W, Schultz TW (1999) A note of caution to users of ECOSAR. Water Qual Res J Canada 34: 179-182.

Kar S, Roy K (2010) QSAR modeling of toxicity of diverse organic chemicals to *Daphnia magna* using 2D and 3D descriptors. J Haz Mat 177: 344-351. [DI 10.1016/j.jhazmat.2009.12.038]

Kenaga EE (1982) Predictability of chronic toxicity from acute toxicity in fish and aquatic invertebrates. Environ Toxicol Chem 1: 347 - 358

Knauer K, Lampert C, Gonzalez-Valero J (2007) Comparison of *in vitro* and *in vivo* acute fish toxicity in relation to toxicant mode of action. Chemosphere 68: 1435-1441. [DI 10.1016/j.chemosphere.2007.04.009]

Lahive E, Halloran JO, Jansen MAK (2011) Differential sensitivity of four Lemnaceae species to zinc sulphate. Environ Exp Bot 71: 25-33.

Landis WG, Chapman PM (2011) Well past time to stop using NOEL/LOELs. Integr Environ Assess Manage 7(4): vi-viii.

Länge R, Hutchinson TH, Scholz N, Solbe J (1998) Analysis of the ECETOC aquatic toxicity (EAT) database II – Comparison of acute to chronic ratios for various aquatic organisms and chemical substances. Chemosphere 36: 115-127.

Lavado R, Maryoung L, Schlenk D (2011) Hypersalinity Acclimation Increases the Toxicity of the Insecticide Phorate in Coho Salmon (*Oncorhynchus kisutch*). Environ Sci Tech 45:4623-4629.

Lavado R, Rimoldi JM, Schlenk D (2009) Mechanisms of fenthion activation in rainbow trout (*Oncorhynchus mykiss*) acclimated to hypersaline environments. Tox Appl Pharma 235:143-152.

Lee BM, Scott GI (1989) The acute toxicity of 4 chemical mosquito control larvicides and a microbial larvicide (BTI) on the estuarine fish, *Fundulus heteroclitus*. Bull Environ Contaminated Tox43: 827-832.

Lemon GD, Posluszny U, Husband BC (2001) Potential and realized rates of vegetative reproduction in Spirodela polyrhiza, Lemna minor, and Wolffia borealis. Aquat Bot 70: 79-87.

Mastrantonio G, Mack H-G, Omar Della Vedova C (2008) Interpretation of the mechanism of acetylcholinesterase inhibition ability by organophosphorus compounds through a new conformational descriptor. an experimental and theoretical study. J Mol Modeling 14: 813-821. [DI 10.1007/s00894-008-0321-0]

Matthews EJ, Kruhlak NL, Benz RD, Contrera JF, Marchant CA, Yang C (2008) Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for windows software to achieve high-performance, high-confidence, mode of action-based predictions of chemical carcinogenesis in rodents. Toxicol Mech Meth 18: 189-206. [DI10.1080/15376510701857379]

Mazzatorta P, Smiesko M, Lo Piparo E, Benfenati E (2005) QSAR model for predicting pesticide aquatic toxicity. J Chem Inf Modeling 45: 1767-1774 [DI 10.1021/ci0502471]

Miller DH, Jensen KM, Villeneuve DL, Kahl MD, Makynen EA, Durhan EJ, Ankley GT (2007) Linkage of biochemical responses to population-level effects: A case study with vitellogenin in the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 26:521-527.

Mkandawire M, Dudel EG. (2005) Assignment of Lemna gibba L. (duckweed) bioassay for in situ ecotoxicity assessment. Aquat Ecol 39: 151-165.

Mkandawire M, Taubert B, Dudel EG (2005) Resource manipulation in uranium and arsenic attenuation by *Lemna gibba* L. (duckweed) in tailing water of a former uranium mine. Water Air Soil Pollut 166: 83-101.

Murawa D, Warminski K, Sulima J, Orzechowski S (2006) QSAR analysis in studies on the acute toxicity of insecticides to aquatic organisms. Pol J Nat Sci 21: 871-881.

Murphy CA, Rose KA, Rahman MS, Thomas, P (2009) Testing and applying a fish vitellogenesis model to evaluate laboratory and field biomarkers of endocrine disruption in Atlantic croaker (*Micropogonias undulatus*) exposed to hypoxia. Environ Toxicol Chem 28:1288-1303.

Natarajan R, Basak SC (2011) Numerical descriptors for the characterization of chiral compounds and their applications in modeling biological and toxicological activities. Curr Top Med Chem 11: 771-787.

Organization for Economic Cooperation and Development (OECD) (2007) OECD Principles for the Validation of (Q)SARS. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono(2007)2&doclanguage=en

Porcelli C, Boriani E, Roncaglioni A, Chana A, Benfenati, E (2008) Regulatory perspectives in the use and validation of QSAR. A case study: DEMETRA model for Daphnia toxicity. Env Sci Technol 42: 491-496. [DI 10.1021/es071430t]

Praba GO, Velmurugan D (2007) Quantitative structure-activity relationship of some pesticides. Ind J Biochem Biophys 44: 470- 476.

Rand GM (Ed) (1995) Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment, Taylor & Francis, Washington, DC.

Reuschenbach P, Silvani M, Dammann M, Warnecke D, Knacker T (2008) ECOSAR model performance with a large test set of industrial chemicals. Chemosphere 71: 1986-1995. [DI 10.1016/j.chemosphere.2007.12.006]

Roy KK, Dixit A, Saxena AK (2008) An investigation of structurally diverse carbamates for acetylcholinesterase (AChE) inhibition using 3D-QSAR analysis. J Mol Graph Modelling 27: 197-208. [DI 10.1016/j.jmgm.2008.04.006]

Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). Environ Toxicol Chem 16:948–967.

Scholz NL, Hopkins WA (2006) The ecotoxicology of anticholinesterase pesticides: data gaps and research challenges. Environ Toxicol Chem 25: 1185-1186.

Schultz TW (1997) TETRATOX: *Tetrahymena pyriformis* population growth impairment endpoint-A surrogate for fish lethality. Toxicol Methods 7: 289-309.

Scott GI (1982) Physiological effects of chlorine-produced oxidants (CPO), dechlorinated effluents, and trihalomethanes (THMs) on marine invertebrates. In: Water Chlorination: Environmental Impact and Health Effects, R.L. Jolley, W.A. Brungs, J.A. Cotruva, R.B.Cumming, J.S. Mattice, and V.A. Jacobs Eds., Ann Arbor Press, Ann Arbor Science Publ., Inc., Ann Arbor, Mich., Vol. 4, pp. 827-841.

Scott GI, Baughman DS, Trim AH, Dee JC (1987) Lethal and sublethal effects of insecticides commonly found in agricultural runoff to estuarine fish and shellfish. In: Pollution and Physiology of Estuarine Organisms, Ed. by WB Vernberg, A Calabreese, FP Thurberg, FJ Vernberg, Baruch series #17, Univ. of South Carolina Press, Columbia, SC. pp 251-273.

Scott GI, Vernberg, WB (1979) Seasonal effects of chlorine-produced oxidants on the growth, survival, and physiology of the American oyster, *Crassostrea virginica* (Gmelin). In: *Marine Pollution*: *Functional Responses*, W.B. Vernberg, A. Calabreese, F. Thurberg, and F.J. Vernberg, Eds., Academic Press, New York City, N.Y., pp. 415-435.

Senior SA, Madbouly MD, El Massry A-M (2011) QSTR of the toxicity of some organophosphorus compounds by using the quantum chemical and topological descriptors. Chemosphere 85: 7-12 [DI 10.1016/j.chemosphere.2011.06.062]

Shi LM, Fang H, Tong WD, Wu J, Perkins R, Blair RM, Branham WS, Dial SL, Moland CI, Sheehan DM (2001) QSAR models using a large diverse set of estrogens J Chem Inf Comp Sci 41: 186-195 [DI 10.1021/ci000066d]

Slavov S, Gini G, Benfenati E (2008) QSAR trout toxicity models on aromatic pesticides. J Env Sci Health B - Pest Food Contam Agric Wastes 43: 633-637. [DI10.1080/03601230802352658]

Slooff W, Vanoers JAM, de Zwart D (1986) Margins of uncertainty in ecotoxicological hazard assessment. Environ Toxicol Chem 5: 841-852.

Spromberg JA, Birge WJ (2005) Modeling the effects of chronic toxicity on fish populations: The influence of life-history strategies. Environ Toxicol Chem 24: 1532–1540.

Teodorović, I, Knežević, V, Tunić, T, Čučak, M, Lečić, JN, Leovac, A, Tumbas, II (2012) *Myriophyllum aquaticum* versus *Lemna minor*: Sensitivity and recovery potential after exposure to atrazine. Environ Toxicol Chem 31: 417–426 [doi10.1002/etc.748]

Tong W, Lowis DR, Perkins R, Chen Y, Welsh, WJ, Goddette, DW, Heritage, TW, Sheehan, DM (1998) Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. J Chem Inf Comp Sci 38: 669-677 [DI 10.1021/ci980008g]

Toropov AA, Benfenati, E (2006) QSAR models for Daphnia toxicity of pesticides based on combinations of topological parameters of molecular structures. Bioorg Med Chem 14: 2779-2788. [DI 10.1016/j.bmc.2005.11.060]

U.S. Environmental Protection Agency (USEPA) (1985) Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. National Technical Information Services, Springfield, VA, USA, PB85-227049.

USEPA (1995) Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. EPA/600/R-95/136. Technical Report. Washington, DC.

USEPA Scientific Advisory Board (2009) Report of the SAB Subcommittee on Aquatic Life Water Quality Criteria for Contaminants of Emerging Concern. EPA-SAB-09-007.

USEPA (2011) Review of methods for characterizing effects of pesticides and other chemical stressors to aquatic organisms. December 20, 2011.

Valenti TW, Taylor JT, Back JA, King RS, Brooks BW (2011) Influence of drought and total phosphorus on diel pH in wadeable streams: Implications for ecological risk assessment of ionizable contaminants. Integrated Environmental Assessment and Management 7: 636-647.

Wang G, Li Y, Liu X, Wang Y (2009) Understanding the aquatic toxicity of pesticide: structure-activity relationship and molecular descriptors to distinguish the ratings of tvffoxicity. QSAR Comb Sci 28: 1418-1431. [DI 10.1002/qsar.200960050]

Wang, J. Grisle, S. and D.Schlenk (2001) Effects of salinity on aldicarb toxicity to juvenile rainbow trout (O*ncorhynchus mykiss*) and striped bass (*Morone saxatilis x chrysops*). Toxicological Sciences 64:200-207.

Warne MS, Van Dam R (2008) NOEC and LOEC data should no longer be generated or used. Australas J Ecotox 14:1–5.

Williams ES, Berninger JP, Brooks BW (2011) Application of chemical toxicity distributions to ecotoxicology data requirements under REACH. Environ Toxicol Chem 30: 1943-1954.

Wu B, Ford T, Gu JD, Zhang XX, Li AM, Cheng SP (2010) Computational studies of interactions between endocrine disrupting chemicals and androgen receptor of different vertebrate species. Chemosphere. 80(5): 535-541.

Yan D, Jiang X, Xu S, Wang L, Bian Y, Yu G (2008) Quantitative structure-toxicity relationship study of lethal concentration to tadpole (*Bufo vulgaris formosus*) for organophosphorous pesticides. Chemosphere 71: 1809-1815. [DI 10.1016/j.chemosphere.2008.02.033]

Yan DY, Jiang, X, Yu, GF, Zhao, ZH, Bian, YR, Wang, F (2006) Quantitative structure-toxicity relationships of organophosphorous pesticides to fish (*Cyprinus carpio*). Chemosphere 63: 744-750. [DI 10.1016/j.chemosphere.2005.08.032]

Zhao JS, Wang B, Dai ZX, Wang XD, Kong LR, Wang LS (2004) 3D-quantitative structure-activity relationship study of organophosphate compounds. Chin Sci Bull 49: 240-245 [DI 10.1360/03wb0156]

Zvinavashe E, Du T, Griff T, van den Berg HHJ, Soffers AEMF, Vervoort J, Murk AJ, Rietjens IMCM (2009) Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*. Chemosphere 75: 1531-1538. [DI 10.1016/j.chemosphere.2009.01.081]