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

OFFICE OF CHEMICAL
SAFETY AND
POLLUTION PREVENTION

MEMORANDUM

DATE: September 11, 2012

SUBJECT: Transmittal of the Meeting Minutes of the FIFRA SAP Meeting Held June 12-14, 2012 on the Scientific Issues Associated with "Problem Formulation for the Reassessment of Ecological Effects from the Use of Atrazine"

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

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

THRU: Laura Bailey 9/12/12
Executive Secretary, FIFRA SAP
Office of Science Coordination and Policy

Frank Sanders 9/12/12
Director
Office of Science Coordination and Policy

Please find attached to this memorandum the meeting report of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) open meeting held in Arlington, Virginia on June 12-14, 2012. This report addresses a set of scientific issues associated with "Problem Formulation for the Reassessment of Ecological Effects from the Use of Atrazine."

Attachment

cc:
Jim Jones
Louise Wise
Vicki Dellarco
William Jordan
Margie Fehrenbach
Donald Brady
Joan Harrigan-Farrelly
Jack Housenger
Richard Keigwin, Jr.
Keith Matthews
Robert McNally
Lois Rossi
Jess Rowland
Oscar Morales
Douglas Parsons
Enesta Jones
Vanessa Vu
OPP Regulatory Docket
Amy Blankinship
Frank Farrugia
Dana Spatz
Russell Erickson
James Hetrick
Anita Pease
Thomas Steeger
Nelson Thurman

FIFRA SAP Members
Daniel Schlenk, Ph.D. (FIFRA SAP Chair)
Marion Ehrich, Ph.D., DABT, ATS
Stephen Klaine, Ph.D.
Jim McManaman, Ph.D.
Martha Sandy, Ph.D.

FQPA Science Review Board Members
Michelle Boone, Ph.D.
William Effland, Ph.D.
Timothy Ellsworth, Ph.D.
Bernard Engel, Ph.D., P.E.
James Fairchild, M.S.
Anne Fetscher, Ph.D.
Robert Gilliom, M.S.
Thomas La Point, Ph.D.
Kenneth Portier, Ph.D.
Catherine Propper, Ph.D.
John Rodgers, Jr., Ph.D.
Michael Twiss, Ph.D.
Linda Young, Ph.D.

SAP Minutes No. 2012-05

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

Problem Formulation for the Reassessment of Ecological Risks from the Use of Atrazine

June 12-14, 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. The 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. 2012-05

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

Problem Formulation for the Reassessment of Ecological Risks from the Use of Atrazine

June 12-14, 2012
FIFRA Scientific Advisory Panel Meeting
Held at
One Potomac Yard
Arlington, Virginia

Daniel Schlenk, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel

Signature:
Date: Sept 7, 2012

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

Signature:
Date: 9/7/2012.

Panel Roster for the Meeting of the Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) to Consider and Review

Problem Formulation for the Reassessment of Ecological Effects from the Use of Atrazine

June 12-14, 2012

EPA-HQ-OPP-2012-0230 **OPP Docket Tel: 703-305-5805**

FIFRA SAP Chair

Daniel Schlenk, Ph.D.
Professor of Aquatic Ecotoxicology & Environmental Toxicology
Department of Environmental Sciences
University of California, Riverside
Riverside, CA 92521

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: 202-564-8450, Fax: 202-564-8382, E-mail: matten.sharlene@epa.gov

FIFRA Scientific Advisory Panel Members

Marion F. Ehrich, Ph.D., DABT, ATS
Professor of Pharmacology and Toxicology
Department of Biomedical Science and Pathobiology
Virginia-Maryland Regional College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0442

Stephen J. Klaine, Ph.D.
Professor and Director
Clemson University Institute of Environmental Toxicology
Department of Biological Sciences
Pendleton, SC 29670

James L. McManaman, Ph.D.
Professor & Chief, Section of Basic Reproductive Sciences
Departments of Obstetrics & Gynecology, Physiology and Biophysics
University of Colorado-Denver
Aurora, CO 80045

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, CA 94612

FQPA Science Review Board Members

Michelle D. Boone, Ph.D.
Associate Professor
Department of Zoology
Miami University
Oxford, OH 45056

William R. Effland, Ph.D.
Soil Scientist
Resource Assessment Division
USDA-Natural Resources Conservation Service
Beltsville, MD 20705

Timothy R. Ellsworth, Ph.D.
Associate Professor
Department of Natural Resources and Environmental Sciences
University of Illinois
Urbana, IL 61801

Bernard A. Engel, Ph.D., P.E.
Professor & Head
Department of Agricultural & Biological Engineering
Purdue University
West Lafayette, IN 47907-2093

James F. Fairchild, M.S.
Research Aquatic Ecologist
United States Geological Survey
Environmental & Contaminants Research Center
Columbia, MO 65201

Anne E. Fetscher, Ph.D.
Senior Scientist - Biology Department
Southern California Coastal Water Research Project (SCCWRP)
Costa Mesa, CA 92626

Robert J. Gilliom, M.S.
Chief, Pesticide National Synthesis
National Water Quality Assessment Program
United States Geological Survey
Sacramento, CA 95819

Thomas W. La Point, Ph.D.
Professor & Director
Institute of Applied Sciences
University of North Texas
Denton, TX 76203-0559

Kenneth M. Portier, Ph.D.
Program Director, Statistics
American Cancer Society
National Home Office
Atlanta, GA 30303-1002

Catherine R. Propper, Ph.D.
Professor
Department of Biological Sciences
Northern Arizona University
Flagstaff, AZ 86011-5640

John H. Rodgers, Jr., Ph.D.
Professor
Department of Forestry and Natural Resources
Clemson University
Clemson, SC 29634-0310

Michael R. Twiss, Ph.D.
Professor
Department of Biology &
Director, Great Rivers Center
Clarkson University
Potsdam, NY 13699

Linda J. Young, Ph.D.
Professor
Department of Statistics
Institute of Food & Agricultural Sciences
University of Florida
Gainesville, FL 32611

INTRODUCTION

The United States Environmental Protection Agency (the EPA or the Agency) Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed its report of the SAP meeting regarding scientific issues associated with **"Problem Formulation for the Reassessment of Ecological Risks from the Use of Atrazine."** Advance notice of the SAP meeting was published in the *Federal Register* on **April 4, 2012**. The review was conducted in an open panel meeting on **June 12-14, 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-2012-0230**. Daniel Schlenk, Ph.D., chaired the meeting and Sharlene Matten, Ph.D., served as the Designated Federal Official.

The EPA provided an overview of the current state of information on atrazine use, environmental fate (exposure), and ecological effects (toxicity) for assessing the potential ecological risk from the use of atrazine. Steven Bradbury, Ph.D., Director, Office of Pesticide Programs (OPP) and Donald Brady, Ph.D., Director, Environmental Fate and Effects Division (EFED), OPP, provided opening remarks at the meeting. Technical presentations were provided by:

1) Dana Spatz, Chief, Environmental Risk Assessment Branch III, EFED, OPP;
2) Amy Blankinship, M.S., Chemist, EFED, OPP;
3) Frank Farruggia, Ph.D., Biologist, EFED, OPP;
4) Russell Erickson, Ph.D., Mid-Continent Ecology Division, Office of Research Development; and,
5) James Hetrick, Ph.D., Senior Science Advisor, EFED, OPP.

Additional technical clarifications were provided by Anita Pease, Ph.D.; Thomas Steeger, Ph.D.; and, Nelson Thurman, Ph.D., from EFED, OPP, EPA.

In 2006, the EPA initiated a program called registration review to re-evaluate all pesticides on a regular cycle as part of the requirements of the Food Quality Protection Act (FQPA). The program reviews each pesticide active ingredient every 15 years to make sure that as the ability to assess risks to human health and the environment evolves and as policies and practices change, all pesticide products in the marketplace can still be used safely. The EPA will soon be reviewing atrazine as part of registration review.

An important step in the development of a risk assessment is the problem formulation. In a problem formulation, available information, including stressor sources and characteristics, exposure, ecological effects on plants and animals (e.g., amphibians, fish, invertebrates, birds, and mammals), and characteristics of the ecosystem(s), is used to define assessment endpoints and to develop a preliminary understanding of potential risks (i.e., develop a risk hypothesis and conceptual model) associated with the use of a pesticide. The problem formulation also serves as an opportunity to identify missing information/uncertainties that may limit the assessment and any assumptions that may be made in the absence of such data. The EPA asked the SAP to peer-review and to provide scientific advice on three specific components of the problem formulation: Topic A - Evaluation of amphibian toxicity data for atrazine, Topic B - Methodology for determining the level of conern (LOC) for atrazine based on aquatic plant communities, and 3) Topic C - Method for comparing monitoring data to the aquatic plant community CE-LOC for atrazine.

The EPA has held four previous SAP meetings on scientific issues associated with the environmental fate and ecological risk assessment of atrazine, specifically, in June 2003, October 2007, December 2007, and May 2009. Each of these meetings reflected the EPA's advancements in understanding the environmental fate and ecological effects of atrazine. The June 12-14, 2012 SAP meeting is the latest in this series. What follows is a brief summary of the previous SAP meetings and the EPA's responses to the SAP's recommendations made at each meeting. This information will provide context for the scientific issues discussed at the June SAP meeting.

The Interim Reregistration Eligibility Decision (IRED) presented the results of the atrazine ecological risk assessment that identified the potential for community and population risk to sensitive aquatic species. Information in the revised IRED was based in part on the review and recommendations of the SAP that met in June 2003 to discuss the potential developmental effects of atrazine on amphibians. At that meeting, the Panel concurred with the EPA's analysis that there was sufficient evidence to formulate a hypothesis that atrazine exposure may impact gonadal development in amphibians, but that there were insufficient data at that time to confirm or refute the hypothesis. This led EPA to seek additional data through a data call-in (DCI) to reduce uncertainties regarding potential risk to amphibians.

In October 2007, the EPA convened a second SAP meeting to evaluate available data on atrazine effects on gonadal development in amphibians. The SAP reviewed the document entitled "*White Paper on the Potential for Atrazine to Affect Amphibian Gonadal Development*" and concurred with the EPA that atrazine does not consistently affect amphibian gonadal development. Although the 2003 SAP indicated that African clawed frog (*Xenopus laevis*) was an appropriate test species given the extent to which the animal is used in amphibian developmental studies, the 2007 Panel concluded that a major uncertainty in the registrant data was the use of ***X. laevis*** as the test organism and the Panel recommended that additional studies were warranted on North American frog species. The SAP acknowledged that there was uncertainty whether study methods for North American species were sufficiently developed or vetted to yield consistent results. Following the October 2007 SAP meeting, the EPA determined that it was reasonable to reject the hypothesis that atrazine exposure can affect gonadal development. Consistent with the recommendations from the 2003 SAP, the Agency also determined, that given the absence of consistent effects and inability to reproduce effects used to support the hypothesis that atrazine affects amphibian development, there was no compelling reason to pursue additional testing with regard to the potential effects of atrazine on amphibian gonadal development. However, the Agency acknowledged that it would continue to monitor research on this subject as it becomes available.

Also as a condition of the 2003 re-registration, the atrazine registrants were required to develop a monitoring program to determine the extent to which atrazine concentrations associated with corn, sorghum, and sugarcane production may be exceeding levels that could cause effects to aquatic plant communities. Forty watersheds representing high atrazine use locations vulnerable to atrazine runoff were selected for monitoring using a stratified, random statistical survey design. Sampling within these watersheds began in 2004 and is ongoing in selected watersheds. There are an additional 25 sites where monitoring began in 2010 to refine the approach for identifying vulnerable watersheds. The EPA is evaluating the results of the atrazine monitoring program, also in part, to identify the characteristics of those watersheds that resulted in atrazine exposures exceeding the Agency's LOC and to extrapolate those results to other non-monitored locations to determine where atrazine concentrations may exceed the LOC.

In December 2007, the EPA presented to the SAP the use of the Comprehensive Aquatic Systems Model (CASM) as a tool to determine a LOC that relates time variable monitoring data to effects identified in a series of microcosm and mesocosm studies. The SAP recommended that the EPA:

1) Work with the CASM-Atrazine model to make the population time series more realistic,
2) Provide a better validation of this model, and
3) Conduct a more comprehensive sensitivity analysis.

In May 2009, the EPA presented a simpler alternative to the CASM-based approach to relate surface water monitoring data to the microcosm and mesocosm data, called the Plant Assemblage Toxicity Index (PATI). Other issues presented at this meeting included a revised assessment of the microcosm and mesocosm exposure profiles, an update on the ecological monitoring program results, interpretation of the monitoring results with PATI, identification of the watershed factors driving atrazine runoff, and extrapolation of those results to the entire atrazine use area to identify other areas where atrazine exposures may exceed the LOC. The 2009 Panel suggested that both the CASM-Atrazine model (presented by Syngenta) and PATI were suitable assessment tools for atrazine. The PATI model was recommended by the SAP as a generic assessment tool for developing the LOC, while CASM was recommended by the SAP as a site-specific assessment tool because of the need for extensive site-specific data. The Panel noted that a limitation in the CASM model is the lack of understanding of the sensitivity of model predictions with respect to correlations among model parameters. The SAP recommended that the EPA re-evaluate the meso/microcosm data set for study quality and concentration-specific effects, and provided additional citations for meso/microcosm studies to consider including in the assessment. They also recommended using a probabilistic approach to determine the LOC. The SAP concurred with the EPA's incorporation of depth to impervious layer and slope to identify vulnerable watersheds for atrazine runoff as part of the atrazine vulnerability index. They also cautioned the EPA that several watershed factors such as atrazine use intensity and rainfall are temporally dependent and, therefore, should not be considered minimum criteria in the vulnerability index.

PUBLIC COMMENTERS

Oral statements were provided by the following individuals:

1) Kerry Kriger, Ph.D., Founder and Executive Director, Save the Frogs

2) On behalf of Syngenta Crop Protection, LLC (Syngenta):
   (a) Dan Campbell, M.S., Atrazine Team Lead, Syngenta
   (b) Janis McFarland, Ph.D., Head of Regulatory Affairs, Syngenta
   (c) Keith Solomon, Ph.D., Professor Emeritus, School of Environmental Sciences, University of Guelph
   (d) Glen Van Der Kraak, Ph.D., Professor and Associate Dean of Research, University of Guelph
   (e) Richard Brain, Ph.D., Environmental Safety, Syngenta
   (f) Jeff Giddings, Ph.D., Compliance Services International
   (g) Malia Andrus, Ph.D., Project Biological Engineer, Waterborne Environmental, Inc.
   (h) Lenwood Hall, M.S., Program Manager Aquatic Toxicology, University of Maryland
   (i) Mark Hanson, Ph.D., Associate Professor, University of Manitoba
   (j) Paul Hendley, Ph.D., Senior Syngenta Fellow, Syngenta

3) On behalf of the Triazine Network:
   (a) Jere White, Executive Director, Kansas Corn and Grain Sorghum Associations
   (b) Laura Knoth, Kentucky Corn Growers Association
   (c) Scott Merritt, Nebraska Corn Growers Association
   (d) Gary Marshall, Missouri Corn Growers Association
   (e) Mark White, Environmental Resources Coalition
   (f) Loren Larson, Caltha LLP

4) Rod Snyder, Director, Public Policy, National Corn Growers Association
5) Scott Slaughter, Center for Regulatory Effectiveness
6) Steven Bruckner, private citizen, McLean, VA
7) Michael Leggett, Ph.D., Senior Director, Environmental Policy, CropLife America
8) Kaitlyn Breyer, private citizen, Rancho Cordova, CA
9) W. Daren Coppock, President & CEO, Agricultural Retailers Association
10) Rick Robinson, Iowa Farm Bureau
11) Roger Johnson, President, National Farmers Union

Written statements were provided by the following individuals:

1) On behalf of Save the Frogs: Kerry Kriger, Ph.D., Founder and Executive Director, and Michael Starkey, Advisory Committee Chairman
2) On behalf of Syngenta: Dan Campbell, M.S., Atrazine Team Lead
3) On behalf of the Pesticides & The Chesapeake Bay Watershed Project: Robert SanGeorge, Project Director
4) On behalf of the Agricultural Retailers Association: Richard Gupton
5) On behalf of the Society for the Study of Amphibians and Reptiles: Betsie Rothermel, Director
6) On behalf of the Minnesota Department of Agriculture: Jim Hines
7) On behalf of CropLife America: Michael Leggett, Ph.D., Senior Director, Environmental Policy

8) On behalf of the Weed Science Society of America: Lee Van Wychen
9) On behalf of the Quaker Concern for Animals: Marian Hussenbux
10) 163 private citizens

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Topic A: Evaluation of the Data on Amphibians

Question 1. *Is the SAP aware of any other laboratory-based or field-based studies not included in this White Paper that should be considered?*

Panel Summary

The Panel agreed that EPA's literature search for papers published prior to the middle of 2011 was very complete. However, since completion of the White Paper several relevant studies have been published which should be considered for risk assessment. The Panel identified 12 published papers that should be considered relevant to environmental exposures to animals. Several of these papers suggested that atrazine disrupted growth and reproduction. Note: The Panel did not evaluate these studies for alignment with the design criteria.

Question 2. EPA identified test design elements that could potentially confound the ability of a study to discern a causal relationship between exposure to atrazine and an effect on amphibians (Section 7.2). Based on consideration of those test design elements, EPA then evaluated the available amphibian data and assigned a classification (*e.g.*, Quantitative, Qualitative (high, medium, and low level of confidence), and Invalid) to each study indicating EPA's confidence in the study's conclusions (Section 7.3 and Appendix C). The confidence in each study was based on an evaluation of the identified test design elements and resulting level of uncertainty in determining a direct causal relationship between atrazine and potential effects to amphibians.

a. Please comment on the completeness of EPA's list of pertinent test design elements. Also, please comment on the degree to which these test design elements, singularly or in combination, would be expected to contribute towards confounding the test results.

Panel Summary

In general, the Panel supported the use of the EPA's test design elements (Section 7.2.b of the White Paper, pgs. 65-66) for use in designing **new** studies for multiple species meant to establish "cause and effect" relationships between the exposure to a specific chemical such as atrazine and the effect of that chemical on an organism. Many of these elements are what statisticians consider as properties of good experimental design prior to the initiation of a study.

When the EPA applied the test design elements retroactively to the 75 amphibian studies in the published literature (36 studies reviewed and presented to the SAP in 2007 and 39 new studies published since the 2007 SAP), only one study on one species was left that met all of the test design elements, the DCI study (Kloas et al. 2009[1]) conducted by the registrant using a strain of *Xenopus laevis.* The earlier

[1] In 2007, Syngenta Crop Protection submitted the final study report, *Atrazine. Response of Larval Xenopus laevis to Atrazine Exposure: Assessment of Metamorphosis and Gonadal Morphology* (Hosmer et al. 2007, MRID 471535-01), for the studies conducted in response to the EPA's 2004 DCI notice. The unpublished study (Hosmer et al. 2007) was later published as Kloas et al. 2009. The Panel will refer to Kloas et al. (2009) when discussing the DCI study.

screening of 36 published studies had the same result, only the DCI study (Kloas et al. 2009) met all of the EPA's test design criteria. The DCI study (Kloas et al. 2009) showed that there were no gonadal effects observed in the strain of *X. laevis* tested at atrazine concentrations 0.01 to 100 µg/L. This one "no effect" *X. laevis* study was used by the EPA to conclude that atrazine has no effect on all amphibians at concentrations less than 100 µg/L. The Panel pointed out that the DCI study may answer the question of whether or not atrazine causes gonadal effects in the strain of *X. laevis* used in the study, but the results of this study are insufficient to make a global conclusion that atrazine has no effect on all amphibian species at concentrations less than 100 µg/L. This was the same conclusion reached by the 2007 SAP (SAP 2008). The EPA's sole reliance on the DCI study (Kloas et al. 2009) using *X. laevis* to conclude that there are no effects on amphibians at concentrations less than 100 µg/L has a strong potential to underestimate the risk of exposure to atrazine across all amphibians.

The Panel disagreed with the EPA's strict application of the test design elements to the published literature that resulted in the elimination of all, but one, of the 75 published studies (the exception was Kloas et al. 2009, the published form of the DCI study) from consideration of potential adverse exposure outcomes. In the view of the Panel, the test design elements should not be applied so strictly to the published literature as to fully disqualify all studies that do not meet **all** of these criteria. One of the steps in determining a cause-effect relationship between atrazine and endpoints at metamorphosis or on reproductive measures requires a randomized experiment in which one or more factors are manipulated individually or in combination, and compared to an appropriate control(s). Although it is completely appropriate to evaluate studies on the basis of their strengths and short-comings, the current approach of selecting and eliminating already published studies errs on the side of failing to detect an effect of atrazine if one exists. In the Panel's analysis, the strict application of the test design elements was flawed and many of the test design elements should be relaxed for review of the published literature. The Panel agreed that while laboratory and field studies have deficiencies, they were concerned that the Agency did not utilize any of these studies to evaluate the effects of atrazine on amphibians. The 2007 SAP expressed the identical concern.

The EPA performed a qualitative evaluation of the 75 peer-reviewed publications. These studies indicated a range of effects from no effect to effects on survival, growth, time and size at metamorphosis, terrestrial performance, hormone levels, increased hermaphroditism, sex ratios, and reproduction. These data suggest that atrazine can affect amphibians at low environmental concentrations, although there is variability in the atrazine concentrations that elicited effects among species. The Panel recommended that most of the studies labeled by the EPA as qualitative or invalid should be considered explicitly in a weight of evidence review of the published studies (a meta-analysis). Particular attention should be given to atrazine concentrations that impact reproductive endpoints, which appear sensitive to atrazine in some populations and species, and might also impact immune responses, which have not been adequately assessed.

The Panel recommended that a "weight of evidence" review of the published studies and new laboratory and cosm studies on more species be conducted before a conclusion can be reached on whether or not atrazine effects amphibians. As did the 2003 and 2007 SAPs, this Panel questioned the suitability of *X. laevis* as a surrogate to represent North American species and recommended that at least three North American species (e.g., a ranid (e.g., *Lithobates [=Rana] pipiens*), a hylid (e.g., *Hyla versicolor*), and a bufonid (e.g., *Anaryxus* [=*Bufo*] *americanus*) from three populations across the U.S. or the species' range, using multiple clutches per population (at least three) should be tested. The Panel suggested that parallel studies be conducted in multiple laboratories using the same test conditions to compare the results of multiple studies to increase the confidence in the conclusions.

b. Please comment on EPA's conclusions about the level of confidence placed on each study's results.

Panel Summary

The Panel agreed that most (though not all) papers deemed "invalid" were not of sufficient quality to capture reliable results. With the exception of these papers, the Panel did not agree with the EPA's conclusions about the level of confidence placed on each study's results by retroactively applying the test design elements to evaluate the "acceptability" of a study in the published literature. See the detailed response to charge question 2a.

The Panel was very concerned with the potential for atrazine to be linked to endocrine disruption in amphibians (e.g., Rohr and McCoy 2010, Tillett et al. 2010, Hayes et al. 2011). These data appear to be largely discounted in the current assessment. The Panel recommended that a high priority be given to planning and conducting future studies to evaluate whether atrazine causes endocrine disruption resulting in adverse effects to survival, development, and reproduction (an adverse-outcome pathway). These studies should be conducted in parallel fashion in multiple independent laboratories to ensure confidence in the results and their interpretation.

Question 3. After evaluating all the available amphibian studies, one study was found to have accounted for all the identified test design elements (Question #2) and determined to be suitable for quantitative use in risk assessment for the endpoints of survival, growth and development (Section 7.3 and Appendix C). T his study was required by an EPA Data Call-In (DCI) Notice following the recommendations from the 2003 SAP on atrazine and amphibians. The resulting study examined the effects of atrazine on *Xenopus laevis* at concentrations of 0.01 to 100 µg/L at two different laboratories. Based on the 2007 SAP, the conclusion was, and there was agreement by the Panel, that the data from this study were robust and sufficient to conclude that exposure to atrazine at concentrations ranging from 0.01 to 100 µg/L had no effect on *X. laevis* development (which included survival, growth, metamorphosis and sexual development).

a. Please comment on whether any new information has become available that leads to a different conclusion from the one which EPA reached in that the results of the DCI study were adequate to evaluate potential effects of atrazine exposure to amphibians.

Panel Summary

Aside from Paetow *et al.* 2012, which found a shift in growth rate in *Lithobates pipiens*, the Panel found no other amphibian studies aside from those listed in the White Paper that have tested effects of atrazine on amphibian growth, reproduction and/or survivorship.

The Panel disagreed with the EPA's characterization (in the charge question) that *"the results of DCI study were adequate to evaluate potential effects of atrazine exposure to amphibians."* Lack of negative effects on *X. laevis* at atrazine concentrations less than 100 µg/L in the DCI study is insufficient evidence to make this same conclusion across all amphibian species and has the potential to underestimate the risk of exposure to atrazine across all amphibians. See the Panel's detailed discussion in response to charge question 2. Note: The 2007 SAP made similar points (see report, SAP 2008).

b. *If such information is now available, please comment on how a threshold determination (a concentration that is expected to cause no effect) may be accomplished using the identified studies.*

Panel Summary

The Panel examined all of the data from Figs. 7, 8, 9 in the White Paper and found that atrazine caused effects on metamorphosis, growth, and sexual development at 1 ppb and above. These summary results suggest that some amphibian species are more sensitive than primary producers. The Panel recommended that determination of the pervasiveness or likelihood of effects at these concentrations and below be considered an important research priority. In support of this risk being significant is the data shown in Fig. 10 in the White Paper, where there are more studies that demonstrate effects than do not (Fisher Exact P<0.01), and solvent does not appear to influence whether or not an effect was found. As such, the weight of solvent used in evaluating confidence in the studies should be reduced and studies that were considered invalid at low solvent levels, e.g., Gunderson et al. (2011), should be re-evaluated to allow for greater confidence in the results. The Panel provided other studies that should be considered in its response to charge question 1.

Question 4. After evaluation of the available amphibian toxicity data, EPA concluded that the DCI study mentioned in Question #3 was appropriate for quantitative use in a risk assessment for survival, growth and development. While the 2007 SAP Panel agreed that atrazine appeared to have no effect on *X. laevis* development at atrazine concentrations ranging from 0.01 to 100 µg/L, they expressed concerns about the suitability of *X. laevis* as a surrogate for native species. Review of the available toxicity data utilizing indigenous species suggests that suitable protocols, including adequate husbandry methods in particular, that would enable EPA to quantify a toxicity endpoint representative of a clear and consistent response from atrazine for native species, may not exist. *Please comment on whether there are suitable methods for testing native amphibians with particular regard to husbandry and laboratory culturing conditions, consistent with the design elements recommended by the 2003 SAP.*

Panel Summary

The Panel agreed with the 2003 and 2007 SAPs that extrapolating results from *X. laevis* to North American species may underestimate exposure risk and recommended that the Agency proceed with further studies conducted with at least three North American species (e.g., a ranid (e.g., Lithobates [=*Rana*] *pipiens*), a hylid (e.g., *Hyla versicolor*), and a bufonid (e.g., *Anaxyrus* [=*Bufo*] *americanus*)) from three populations across the U.S. or the species' range, using multiple clutches per population (at least three) should be tested to provide sufficient evidence on whether atrazine causes or does not cause adverse effects to amphibians.

Although amphibians are not used in routine toxicity testing, and therefore, strict standards for culturing and testing have not been developed, the Panel stated that several laboratories throughout North America have successfully (and sometimes commercially) cultured native species. The Panel encouraged the EPA to proceed with testing North American amphibian species and suggested a number of researchers (laboratories) who would be useful resources on how the EPA's test design elements could be adapted to current culturing practices to meet species-specific husbandry needs. Individuals who have experience with native North American species include: Helbing (University of Victoria; experience with *Lithobates catesbeiana and Pseudacris regilla*), Trudeau (University of Ottawa; experience with *Lithobates sylvaticus and L. pipiens*), Relyea (University of Pittsburgh; experience with

Ranidae: *L. pipiens, L. clamitans, L. sylvatica, L. catesbeiana, L. cascadae*; Bufonidae: *Anaxyrus americanus, B. boreas; and Hylidae: Hyla versicolor, Pseudacris crucifer*), and Boone (Miami University, Ohio; Ranidae: *L. pipiens, L. clamitans, L.catesbeiana*; Hylidae: *H. chrysoscelis, H. versicolor, Acris crepitans, P. regilla*).

Mesocosm studies (with 3-5 replicates) are routinely used in amphibian ecology and ecotoxicology, and yield high power, often >0.8, depending on the endpoint and effects size. Typically-used methods are described in Semlitsch & Boone (2009). The Panel agreed with the 2003 SAP's recommendation that it is necessary to conduct field mesocosm studies, as effects of atrazine can change between the laboratory and field (SAP 2003). Cosm studies should be conducted under more natural field conditions to examine how natural factors influence the impact of atrazine on amphibians. Such studies allow for the determination of cause-effect relationships.

The Panel recommended that all test design criteria intended to be used in the animal husbandry community undergo further peer review. The animal husbandry community needs to understand how to design their laboratory or mesocosm studies to meet these criteria. The Panel suggested that this information be published in an appropriate peer-reviewed journal and/or EPA web-sites, and distributed through the amphibian *ecotox list-serv*, as well as disseminated to researchers conducting studies with atrazine and other pesticides.

Question 5. A number of studies report the potential for atrazine to modify immune function and infection susceptibility in amphibians (Appendix C). EPA believes the research on these different hypotheses does not provide sufficient data to establish causal linkages among different levels of biological organization to result in adverse effects. Therefore, EPA concluded that a mode of action or adverse outcome pathway leading to effects on amphibian survival, growth or development cannot be established at this time.

a. Please comment on whether the data in the existing database reasonably supports the hypotheses, or demonstrates that atrazine affects immune function and/or infection susceptibility leading to adverse effects on survival, growth or development; i.e., are there sufficient data to establish an adverse outcome pathway for atrazine effects on immune function? Please provide a rationale for the Panel's position and discuss the associated strengths and weakness with the data supporting the rationale.

Panel Summary

There are relatively few studies to date examining atrazine's effect on immune responses or susceptibility to disease, pathogens, or parasites, making it difficult to determine the importance of these effects. For this reason, the Panel agreed with the EPA's conclusion that existing data are insufficient to support or refute hypotheses that atrazine adversely affects immune functions and infection susceptibility leading to adverse effects on survival, growth or development (i.e., an adverse outcome pathway for atrazine cannot be established). The Panel considered the available data to be adequate to conclude that there "could" be adverse outcomes and sufficient to justify further investigation into atrazine and its potential impact on disease susceptibility and immune function leading to adverse effects.

b. If the Panel concludes that the existing data are sufficient to formulate hypotheses that atrazine adversely affects immune function and infection susceptibility, but are not sufficient to test the hypotheses (refute or confirm), then please comment on specific study

protocols that can be used to test these hypotheses with sufficient rigor to identify effects that can be directly and quantitatively attributed to adverse impacts on amphibian reproduction, growth and/or survival.

Panel Summary

The Panel referred to its response to charge question 5a. The available data suggest that atrazine could increase susceptibility of amphibians to disease and affect immune function. Critical endpoints and protocols should be developed by the scientific community (e.g., immunologists). The Panel recommended that research in this area be considered a high priority to address questions concerning potential atrazine adverse effects on immune function and infection susceptibility. The Panel offered several recommendations to evaluate potential atrazine effects on disease susceptibility.

Topic B: The Method for Determining the Atrazine Level of Concern for Aquatic Plant Communities

Question 6. The cosms were comprised of natural communities of periphyton/phytoplankton; in some cases, vascular plants, invertebrates and vertebrates present in those communities were included in the study (Chapter I, Section 6.1.4). These sources were generally described as streams, lakes, reservoirs, and springs, and are considered to be representative of the structure and function of aquatic plant communities in such water bodies. *Given the diversity of sources and the described communities, please comment on the extent to which these cosm studies taken together provide useful and reasonable physical models of the natural aquatic plant communities exposed to atrazine in the U.S.*

Panel Summary

The Panel stated that it was critical to have a common set of criteria to score each cosm study since the calculation of the level of concern for effects on aquatic plant communities (CE-LOC) using the PATI model is highly sensitive to which cosm studies were included in the analysis. This situation does not currently exist and needs to be resolved before a CE-LOC can be determined.

The Panel provided the following conclusions and recommendations to advance the analysis of existing cosm studies and design of new studies.

Re-evaluate and re-score existing cosm studies with purported effects identified at atrazine concentrations less than 30 µg/L. Most of the existing cosm studies with purported effects identified at atrazine concentrations less than 30 µg/L have weaknesses in their design that render interpreting their results and scoring them for "effects" or "no effect" difficult and subjective. The Panel recommended that the cosm studies at concentrations less than 30 µg/L (Appendix D, White Paper) be re-evaluated and re-scored using a common set of scoring criteria, e.g., Giddings (2012). The Panel's re-evaluation of this subset of cosm studies identified 11 cosm studies mis-scored by the EPA as having "effects," when they should have been re-scored as having "no effect" (i.e., less than 30 µg/L, see Table 1). The Panel suggested that these studies be re-scored and/or dropped from the dataset. The 2009 SAP made a similar recommendation. The Panel noted that EPA would not have to re-evaluate this subset of cosm studies if they adopted the scoring of Giddings (2012). This would be appropriate since the rescoring by the Panel of this subset of studies agreed with that of Giddings (2012). The outcome of this entire exercise would be to determine which studies showed effects at atrazine

concentrations less than less than 30 µg/L that were truly statistically significant and ecologically relevant. Many of these studies were conducted at the University of Kansas and do not actually contain all of the information necessary to score them since much of the text merely alludes to the original discussion in students' Masters Degree theses or other papers. Studies at atrazine concentrations greater than or equal to 30 µg/L do not need to be re-scored because there is agreement on the scores. The Panel would like to see the CE-LOC re-calculated with the re-scored studies to determine the relative effect on the CE-LOC.

Additional focused cosm studies are needed to specifically address the durations of exposure expected at environmentally relevant concentrations. The Panel overwhelmingly agreed and strongly recommended that additional focused cosm studies are warranted to specifically address the durations of exposure expected at environmentally relevant concentrations. Results from these studies would be used to fill in the data gaps that currently exist for determining whether there are statistically significant and ecologically relevant atrazine effects at the 4 to 7 µg/L level (proposed as the CE-LOC in the Agency's White Paper) and also at higher concentration levels.

The Panel suggested that a cosm study be developed that meets the criteria for a high quality, quantitative study (described above). The design of the cosm study should be based on the exposure scenario for a 2^{nd}-3^{rd} order stream (i.e., lotic system) with an emphasis on periphyton, and using a set of primary producer species representing a broad range of sensitivities to anthropogenic stress. The same set of measurement endpoints used in the Syngenta field reconnaissance studies conducted by Andrus et al. in Missouri and Illinois presented to the Panel would be a good start in the design of the study (see Andrus et al. presentation in the public docket, EPA-HQ-OPP-2012-0230). In addition, the study should be conducted using a replicated experimental design with concentrations that bracket the 4 to 7 µg/L proposed CE-LOC (i.e., 0, 4, 8, 16 µg/L) and extend to concentrations which are more likely to have statistically-significant ecological effects (i.e., 32, 64, and128 µg/L). Currently, there are no lotic studies at this range of concentrations in the database.

Proceed with evaluation of atrazine as the cosm data are refined and improved. While recommending further cosm studies, the Panel also stated that the Agency should proceed with the evaluation of atrazine as plans are developed for implementing further studies. Specifically, the re-evaluation of cosm studies may lead to a less or greater need for the new studies than presently perceived.

Question 7. The Aquatic Plant Community CE-LOC methodology for atrazine is a four stage process that uses single-species plant toxicity data and cosm studies to discern atrazine concentrations and exposure durations that may cause adverse effects on aquatic plant communities. As a result, a CE-LOC for atrazine is developed which, together with monitoring data, can be used to identify watersheds where concentrations may result in adverse effects to aquatic plant community structure, function, and/or productivity. *Please comment on the methodology EPA has used to derive the atrazine CE-LOC for aquatic plant communities, and in particular on EPA's characterization of the uncertainties and assumptions in this methodology (Chapter IV, Sections 13 & 14).*

Panel Summary

The Panel was very impressed with the amount of work and consideration put into these calculations and estimates. The Panel agreed that the four stages of the aquatic plant community LOC methodology for atrazine were logical. Nevertheless, the Panel expressed genuine concerns with aspects of the LOC methodology, notably the structure and assumptions of PATI and scoring the cosm data. As a result, the Panel expressed minimal confidence in the calculated CE-LOC.

The Panel presented several methods for characterizing this uncertainty. One general approach would be to use a bootstrap method to quantify the uncertainty inherent in the data, e.g., sampling with replacement "B" times, constructing estimates for each of the "B" samples and assessing the variability of the estimates. Another option would be to use a series of sequential random samplings without replacement of the cosm data to examine if there is a convergence in CE-LOC estimates with increasing cosm data. A third approach would be to group the data by exposure duration (e.g., estimate the CE-LOC using the cosm data with exposure durations of 20 days or less), and then sequentially add subsequent cosm data with increasing exposure durations to evaluate the influence of cosm exposure durations on the CE-LOC estimate. The Panel noted that the latter two methods are not as statistically rigorous as the bootstrap method.

The Panel was concerned that the single-species toxicity tests (Stage 2 of the four-stage process, Fig. 15, p. 100 of the White Paper) were based largely on species typically observed to be "tolerant" to various forms of anthropogenic stress, which could result in underestimating the effects of atrazine in natural plant communities. Therefore, the Panel recommended that a broader range of species sensitivities to atrazine be included in such tests to better represent the potential effects to a full of range of primary producer species in the system and maintenance of ecosystem integrity.

The Panel also had concerns about the degree of intraspecific variation in some of the SGR EC_{50} values used in deriving the PATI relationship. The Panel recommended that study-acceptability criteria be incorporated in the screening process to ensure that some meaningful portion of the studies used in the development of the CE-LOC include species known to occur in reference sites where atrazine is absent, as opposed to focusing largely on potentially tolerant species, or species common in generally disturbed areas, that are typically cultured in laboratory settings. Studies should provide support that the strains used have not been exposed to atrazine (i.e., they should come from previously unexposed, wild populations) in order to establish the representativeness of the observed toxicity response for a given species. If these additional criteria are not met by the studies used, then the CE-LOC should be qualified in a way that clearly reflects that it is not known whether the CE-LOC represents the full range of primary producer sensitivity to atrazine.

Question 8. The 2009 SAP recommended using an effects index or concentration metric, rather than categorical LOC thresholds in order to take advantage of data from Syngenta's Atrazine Ecological Exposure Monitoring Program (AEEMP). At that time the LOC threshold for atrazine effects to plant communities was established at 10 µg/L for a 60-day rolling average. The current analysis using the Plant Assemblage Toxicity Index (PATI) indicates the CE-LOC can range from 4 to 7 µg/L (Chapter IV, Section 14.3 & 14.4). *Please comment on this CE-LOC and whether it reasonably represents a range below which permanent or irreversible change in aquatic plant community structure, function, and/or productivity due to atrazine exposure would not be expected.*

Panel Summary

The majority of the Panel indicated that the 60-day rolling average of 4 to 7 µg/L CE-LOC proposed by the Agency represents a level below which no permanent or irreversible change in aquatic plant community structure and/or productivity will occur. Having said this, the Panel could not validate the 4 to 7 µg/L CE-LOC due to a number of concerns with the LOC methodology previously described in responses to charge questions 6 and 7. In brief, the Panel had concerns with the selection process of the final cosm dataset. Furthermore, each step in the multi-step LOC methodology is associated with inherent error, and then propagated along each step so that the accumulated error in the CE-LOC will likely be quite large. As a result, the Panel expressed minimal confidence in the calculated CE-LOC.

Question 9. Based on previous analyses of the available ecotoxicity data, EPA concluded for atrazine that the level of concern for effects on aquatic plant communities (CE-LOC) was lower than the atrazine concentrations observed to produce significant direct or indirect effects on invertebrates, fish and amphibians. Given the current analysis of the ecotoxicity data (Chapter I, Section 6) and the Aquatic Plant Community LOC methodology, EPA continues to believe the original conclusion still holds true. *Please comment on how well the available database supports EPA's conclusion that the CE-LOC is lower than exposures that result in significant effects on the growth, survival and reproduction of aquatic animals.*

Panel Summary

The Panel stated that without further testing of the effects of long-term, low concentration exposures of atrazine and its degradation products, EPA's conclusion (stated in the charge question) that the *"CE-LOC is lower than exposures that result in significant effects on the growth, survival and reproduction of aquatic animal"* cannot be supported. The Panel noted that a regression design could cover a range of environmentally relevant exposures without being constrained by an estimated CE-LOC. As of this date, there are no published data indicating that environmentally relevant concentrations of atrazine at or below 4 to 7 µg/L cause mortality to invertebrates, fish or amphibians (e.g., see analysis by Rohr and McCoy 2010), although there are some published studies that suggest that atrazine exposures could lead to effects on reproduction, behavior, and development.

The EPA's acceptance criteria for cosm studies (Appendix D, p. 3) indicated that tests must involve administration of atrazine only and not mixtures or multi-active ingredients. The Panel pointed out that isolation of atrazine for toxicity testing and cosm studies likely render direct assessment of atrazine effects much more straightforward in the laboratory. However, it would be flawed in terms of how atrazine affects growth, survival, and reproduction of aquatic organisms under "realistic conditions" because atrazine will be only one of the many stressors in the environment which could act additively or synergistically with atrazine to heighten the effective toxicity of a given measured concentration of atrazine. Therefore, likely interactions among environmental stressors with atrazine should be considered, or at least not eliminated from the analysis when available in the published literature, in the process of establishing meaningful effects levels.

Furthermore, the Panel noted that the EPA's rationale for not including atrazine degradates in the analysis overlooked the possibility that a given measured concentration of atrazine in the presence of its (unmeasured) degradates may be more toxic than that same concentration of atrazine in the absence of degradates. To the extent that atrazine in combination with its degradates results in toxic effects exceeding those that would be anticipated from monitoring of, and accounting for, atrazine alone, omission of degradates from consideration of exposure may result in underestimating risk.

Topic C: Method for Comparing Monitoring Data to the Aquatic Plant Community LOC for Atrazine

Question 10. *Please comment on the strengths and limitations of EPA's development and use of bias factors (Chapter V, Section 16.1) for addressing uncertainties in monitoring data.*

Panel Summary

The Panel commented on the strengths and weaknesses with the use of bias factors to address uncertainties in the monitoring data and suggested a number of clarifications and refinements that should improve both the acceptability and understanding of the bias factor approach. These included, but were not limited to, the rationale for the selection of the data used in the analysis, examples and details on in-filling, use of these data in the regression analysis and choosing representative monitoring sites. Many of these were previously discussed during the 2011 SAP meeting (SAP 2011).

Question 11. Prediction of bias factors is dependent on the selection of an appropriate model. EPA illustrated (Chapter V, Section 16.1) both categorical and regression methods for prediction of bias factors based solely on the number of samples taken in the 2nd and 3rd quarters of the year (April 1st to September 30th).

a. Please comment on EPA's prediction of bias factors from monitoring data using categorical or regression method approaches.

Pannel Summary

The Agency acknowledged that the bias factor prediction methods are preliminary estimates based on only seven site year combinations. The Panel emphasized the importance of expanding the data to provide a more representative assessment for the regions of interest. Albeit preliminary estimates, the Panel enumerated several issues with the regression (three) and categorical (one) approaches used to select bias factors. The Panel recommended that bias factors be estimated for other sampling intervals, which could be integrated into the bias factor prediction model.

b. Please comment on any additional methods for estimating bias factors that would be useful in this situation.

Panel Summary

The Panel discussed several strategies for selecting which sites to use for the further development of the bias factor approach. One approach discussed was to consider all of the correlated information that is available for estimating stream atrazine concentrations, e.g., use of some type of regressor like the stream flow level/flow rate measurements, PRZM-hybrid approach, Lerch approach, Rinaldo approach, Corn Belt (CB) WARP, to estimate the concentration rather than simply linear interpolation. Another option would be to use multiple approaches to predict the 60-day rolling average (e.g., PRZM-Hybird, CB WARP). In this regard, these predictions may actually be more accurate (less uncertainty) than the linear interpolation approach. Another suggestion was to use sampling interval bias factors directly. Finally, many on the Panel disagreed with the EPA's assumption that the grab sample estimated 60-day average is the true 60-day rolling average and recommended a daily composite sample approach.

Question 12. EPA illustrated (Chapter V, Section 16.1) both categorical and regression methods for estimation of bias factors as a function of the sampling frequency of monitoring data. Step-wise regression analysis indicates that watershed size and average flow rate in the 2nd and 3rd quarters of the year are not significant variables for prediction of bias factors. However, the number of samples in the 2nd and 3rd quarters of the year was found to be a significant variable, accounting for 46% of the variation in the bias factor. *What other variables, if any, should be considered in the prediction of bias factors?*

Panel Summary

The EPA performed a preliminary analysis based on seven site year combinations for bias factor estimation. The Panel indicated that the next step is to increase the data set to produce a regionalized (e.g., Corn Belt) estimation of bias factors (see also response to charge question 10). A comprehensive set of bias factors based on as much of the available monitoring data as possible should be developed prior to attempting to predict bias factors as a function of physically based characteristics (e.g., sampling scheme, watershed area, topography, land use classifications, hydrologic response indices, climatic factors, etc.). Several panel members commented that a number of the general points made in response to this charge question were also discussed at the 2011 SAP meeting (see SAP 2011). The Panel provided specific recommendations with respect to composite sampling, methods for in-filling data, and watershed classification.

Question 13. EPA examined (Chapter V, Section 16.1) the performance of various regression equations to assess the failure percentage for identification of monitoring site-years with true maximum 60-day average concentrations exceeding the CE-LOC for atrazine. This analysis showed that application of a bias factor, based on sample number during the 2nd and 3rd quarter of the year, substantially reduced the number of sites with underestimation of true maximum 60-day means. *Given the EPA analysis, what other tests, if any, should be conducted to assess the performance of regression models for prediction of bias factors?*

Panel Summary

The Panel recommended that the performance of the final models be addressed by using standard regression techniques and diagnostics to achieve the best fitting model, e.g., goodness-of-fit test. Secondly, regression models should be evaluated with an independent data set that covers the full geographic and hydrographic range of watersheds to which the model is applied. If about 2.5% of the bias factors are above the upper 95% prediction limit, then the model is performing well. The "leave-one-out" cross validation method assures the fit of the model if the data set is not large enough to allow a separate testing data set. The Panel indicated that combining several conservative processes to establish limits on the true maximum 60-day mean concentrations will likely cause the final limit to be unrealistically conservative. First, the PATI-LOC was derived based on cosm studies which did not adequately evaluate the recovery from atrazine exposures. Second, the cosms effects/no effects were very conservatively rated by the EPA (mis-scored according to the Panel as discussed in charge question 6) and the PATI LOC is extremely sensitive to the cosm ratings. Third, the bias factor estimates employ a 5th percentile cutoff which results in a very conservative estimate of 60-day rolling average concentrations and considers the frequency of occurrence of false positives.

Question 14. An important consideration for bias factor prediction is the ability to identify monitoring sites that potentially exceed the CE-LOC for atrazine. EPA provided an example (Chapter V, Section 16.2 and 16.3) of the use of a log-linear regression equation to estimate bias factors from the USGS National Water Quality Assessment Program NAWQA data and identification of monitoring site-years exceeding the CE-LOC for atrazine. This analysis identified sites in both the Midwestern United States as well as outside the major corn growing areas with atrazine concentrations potentially above the CE-LOC. *What are the strengths and limitations of using a single regression model for prediction of spatially and temporally explicit bias factors for a nationally-distributed monitoring data set*?

Panel Summary

The Panel indicated that a single regression model is a simple, uniform and transparent approach to estimating bias factors from using a wide range of existing and potentially future data sets to estimate site and year specific estimates of the maximum 60-day average for comparison to the CE-LOC. However, the model is limited by the representativeness and completeness of the model development data set and the quality of the final model to represent those conditions. The greatest limitation of the approach is that most stream miles for which estimates are needed have no monitoring data that will meet the minimum requirements for making an initial sample estimate to be bias-corrected. With more monitoring sites (and site years) and watershed characteristics, the Panel stated that the log-linear model will likely be inadequate to describe the majority of the variability in the data. More complex models, such as the multivariate regression tree model (De'ath 2002, Prasad et al. 2006), might be explored. This approach can facilitate both estimation of bias factors and identification of conditions under which the uncertainty of the resulting bias factor estimate is too high to reliably use this approach.

The Panel also suggested that if one were to examine the probability of exceeding the CE-LOC, then the entire range of methods from sample estimates based on intensive data, to bias-corrected sparser data, to regression-based multivariate extrapolation models, could all be used and judged by the same standard. All of these methods could be part of a set of methods that provide a step-wise process for identifying and dealing with streams of potential concern.

DETAILED RESPONSES TO CHARGE QUESTIONS

Question 1. There are an increasing number of literature studies examining whether there is a causal relationship between atrazine exposure and abnormal sexual development, metamorphosis and growth, and/or immune response in amphibians (Chapter II of the White Paper). In order to determine the degree to which these studies are appropriate for incorporation into risk assessment, EPA evaluated the data within the context of considering the nature and degree of the uncertainties in each study. *Is the SAP aware of any other laboratory-based or field-based studies not included in this White Paper that should be considered?*

Panel Response

The Panel agreed that EPA's literature search for papers published prior to the middle of 2011 was very complete. However, since completion of the White Paper several relevant studies have been published which should be considered for risk assessment. The Panel identified 12 published papers that should be considered relevant to environmental exposures to animals. Several of these papers suggested that atrazine disrupted growth and reproduction. Note: The Panel did not evaluate these studies for alignment with the design criteria.

Among these papers, there were a wide range of species and endpoints evaluated. In general, most studies found anywhere from low to high shifts in outcome measures associated with atrazine exposure, including effects on physiological function, shifts in gene expression, hormone levels, growth, gill function, behavior, reproductive function, and hatching success. Several of these papers suggested that atrazine disrupted growth and reproduction. The Panel agreed with the EPA that there should be high certainty in ascribing effects noted in the studies specifically to atrazine. However, the Panel stated that a risk assessment relying on only one (or two) studies to evaluate an entire species, let alone a whole class of animals, was inadequate (see Panel's response to charge question 2).

A summary of the 12 publications recommended by the Panel (not included in EPA's analysis) for further consideration by the EPA is provided below (references provided at the end of the report). These studies are divided into animal groups. The outcomes of these studies demonstrate that there is variation in results across species, doses and design.

1) **Snails.** There were no effects on snail growth in mesocosms treated with atrazine. (Baxter et al. 2011).

2) **Fish.** In a variety of fish species, atrazine exposure impaired the stress response and glucose metabolism (Nascimento et al. 2012), gill function (Paulino et al. 2012a; Paulino et al. 2012b), growth (Corvi et al. 2012), and behavior (Shenoy 2012; Tillitt et al. 2010). The Panel noted that the Tillet et al. (2010) paper was mentioned on page 56 of the White Paper as being one that will be further evaluated for inclusion in the atrazine risk assessment, but it is not mentioned again in the White Paper or in Appendices B and C. This study noted a decrease in egg production resulting from fewer spawning events in fathead minnows exposed to atrazine. Effects were measured at doses as low as 0.5 ppb, and represent a reduction in fitness, making this study of particular importance to the potential for population-level outcomes. Corvi et al. (2012), using a zebrafish model, found no statistical effect of atrazine on gonadal development and no effect on spawning.

3) **Amphibians.** There was one new amphibian study that was not evaluated by the EPA, Paetow et al. (2012). These authors found shifts in growth , but not in infection rate to the amphibian chytrid fungus (Paetow et al. 2012). The Panel was particularly interested in one previously published paper not considered in the White Paper, Bishop et al. (2010). Though the field study involved multiple chemicals, results indicated that atrazine exposure was associated with reduced hatching success (Bishop et al. 2010). The Panel suggested that this study and similar types of studies could be used in a weight of evidence approach to assess the likelihood of atrazine having effects in natural, complex environments.

4) **Mice.** Behavior was also affected in mice (Belloni et al. 2011) suggesting neurological outcomes following exposure to atrazine. These results along with those from previous papers that also demonstrated that there were behavior shifts in amphibians following exposure to relatively low doses of atrazine, point to the possibility of neurological disruption which should be further studied.

5) **Humans.** Although human effects are not related to the charge questions issued for this SAP, results from a relatively recent paper suggested that atrazine exposure was associated with shifts in cycle length in women (Cragin et al. 2011).

Question 2. EPA identified test design elements that could potentially confound the ability of a study to discern a causal relationship between exposure to atrazine and an effect on amphibians (Section 7.2). Based on consideration of those test design elements, EPA then evaluated the available amphibian data and assigned a classification (e.g., Quantitative, Qualitative (high, medium, and low level of confidence), and Invalid) to each study indicating EPA's confidence in the study's conclusions (Section 7.3 and Appendix C). The confidence in each study was based on an evaluation of the identified test design elements and resulting level of uncertainty in determining a direct causal relationship between atrazine and potential effects to amphibians.

a. Please comment on the completeness of EPA's list of pertinent test design elements. Also, please comment on the degree to which these test design elements, singularly or in combination, would be expected to contribute towards confounding the test results.

Panel Response

In general, the Panel supported the use of the EPA's test design elements (Section 7.2.b of the White Paper, pgs. 65-66) for use in designing **new** studies for multiple species meant to establish "cause and effect" relationships between the exposure to a specific chemical such as atrazine and the effect of that chemical on an organism. Many of these elements are what statisticians consider as properties of good experimental design prior to the initiation of a study.

The Panel disagreed with the EPA's strict application of the test design elements to eliminate all of the 75 published studies from consideration of potential adverse exposure outcomes, except for the DCI study on *Xenopus laevis* (Kloas et al. 2009[2]). The DCI study (Kloas et al. 2009) showed that there were

[2] In 2007, Syngenta Crop Protection submitted the final study report, *Atrazine. Response of Larval Xenopus laevis to Atrazine Exposure: Assessment of Metamorphosis and Gonadal Morphology* (Hosmer et al. 2007, MRID 471535-01), for the studies conducted in response to the EPA's 2004 DCI notice. The unpublished study (Hosmer et al. 2007) was later published as Kloas et al. 2009. The Panel will refer to Kloas et al. (2009) when discussing the DCI study.

no gonadal effects observed in the strain of *X. laevis* tested at atrazine concentrations 0.01 to 100 µg/L. This one "no effect" *X. laevis* study was used by the EPA to conclude that atrazine has no effect on all amphibians at concentrations less than 100 µg/L. The Panel pointed out that the DCI study may answer the question of whether or not atrazine causes gonadal effects in the strain of *X. laevis* used in the study, but the results of this study are insufficient to make a global conclusion that atrazine has no effect on all amphibian species at concentrations less than 100 µg/L.

The Panel questioned the suitability of *X. laevis* as surrogate to represent North American species and recommended additional testing in North American species. The 2007 SAP also discussed the limitations of the DCI study including those associated with the strain of *X. laevis* used, flow-through system vs. static system, and extrapolating the results from *X. laevis* to North American species (SAP 2008). The EPA's sole reliance on the DCI study using *X. laevis* to conclude that there are no effects on amphibians at concentrations less than 100 µg/L has a strong potential to underestimate the risk of exposure to atrazine across all amphibians.

One of the steps in determining a cause-effect relationship between atrazine and endpoints at metamorphosis or on reproductive measures requires a randomized experiment in which one or more factors are manipulated individually or in combination, and compared to an appropriate control(s). The Panel concluded that many of the studies designated as "qualitative" met this basic requirement for establishing cause-effect relationships and were peer-reviewed by some of the top-tier journals; however, these studies have not been used to reach conclusions regarding the effect of atrazine on amphibians because they have not met all of the EPA's test design elements. Although it is appropriate to evaluate studies on the basis of their strengths and short-comings, the current approach of selecting and eliminating already published studies errs on the side of failing to detect an effect of atrazine if one exists.

The Panel agreed that while laboratory and field studies have deficiencies, they were concerned that the Agency did not utilize any of these studies to evaluate the effects of atrazine on amphibians. The 2007 SAP expressed the identical concern. In the view of the Panel, the test design elements cannot be applied so strictly to the published literature as to disqualify all studies that do not meet **all** of these criteria. In the Panel's analysis, the EPA's strict application of the test design elements to the published literature was flawed and many of the test design elements should be relaxed for review of the published literature.

The EPA performed a qualitative evaluation of the 75 peer-reviewed publications. Results from these studies indicated a range of effects from no effect to effects on survival, growth, time and size at metamorphosis, terrestrial performance, hormone levels, increased hermaphroditism, sex ratios, and reproduction. These data suggest that atrazine can affect amphibians at low environmental concentrations, although there is variability in the atrazine concentrations that elicited effects among species. The Panel stated that all studies (qualitative and those labeled as invalid by the EPA) should be considered explicitly in the risk assessment as they may provide qualitative support for the hypothesis that atrazine causes effects in many species of amphibians, e.g., reproductive effects, immune responses.

The EPA concluded that "*based on the available data, atrazine did not appear to produce consistent effects on amphibian development and that based on tiered testing approach reviewed by the 2003 SAP, no further testing was needed"* (p. 62 of White Paper). The Panel disagreed with this conclusion and made the following two recommendations.

1) The Panel recommended that a weight of evidence approach (a meta-analysis) be used to review all of the published studies (quantitative, qualitative, and those labeled as invalid by the EPA). All studies should be considered explicitly in the risk assessment.

2) The Panel recommended that at least three North American species (a ranid (e.g., *Lithobates [=Rana] pipiens*), a hylid (e.g., *Hyla versicolor*), and a bufonid (e.g., *Anaryxus* [=*Bufo*] *americanus*) from three populations across the U.S. or the species' range, using multiple clutches per population (at least three) should be tested before a conclusion can be reached that atrazine has no effect on amphibians at concentrations less than 100 µg/L.

Analysis of the EPA's design elements

The Panel agreed that a review of the published literature is necessary to provide context for any individual study among the larger group of published studies. To this extent, the EPA established test design criteria to sort studies into three categories, "quantitative, qualitative, and invalid," to provide a degree of confidence in each individual study. These criteria include the major confounding elements observed in laboratory and field studies that were originally discussed during the 2003 SAP meeting (see SAP 2008). Many of these elements are what statisticians consider as properties of good experimental design prior to the initiation of a study, e.g., sufficient replication, adequate controls, and adequate randomization to avoid bias. Studies conducted in the laboratory or the field should have an auditable QA/QC trail designed to generate data that are unequivocal and provide greater confidence in the conduct of the studies if confounding test design elements are eliminated.

The Panel determined that the EPA's test design elements are very useful to use in designing new studies, but not when they are applied retroactively to the published literature. When the EPA's test design elements were strictly applied to the 75 amphibian studies in the published literature (36 studies reviewed in 2007 and 39 new studies published since 2007), only one study (two identical experiments conducted in two different laboratories) on one species was left that met all of the test design elements, the DCI study conducted by the registrant using a strain of *X. laevis* (Kloas et al. 2009). In 2007, the EPA used test design criteria to screen 36 amphibian studies (from the published literature and those submitted by the registrant), and again, only one study, i.e., the DCI study conducted by the registrant using a strain of *X. laevis* (Kloas et al. 2009), met every one of the test design elements. The DCI study (Kloas et al. 2009) showed that there were no gonadal effects observed in the strain of *X. laevis* tested at atrazine concentrations 0.01 to 100 µg/L. This one "no effect" *X. laevis* study was used by the EPA to conclude that atrazine has no effect on all amphibians at concentrations less than 100 µg/L. The Panel pointed out that the DCI study may answer the question of whether or not atrazine causes gonadal effects in the strain of *X. laevis* used in the study, but the results of this study are insufficient to make a global conclusion that atrazine has no effect on all amphibian species at concentrations less than 100 µg/L.

In many cases, the Panel determined that the strict application of the EPA's test design elements to the published literature was flawed and recommended that less concern be given to some of these elements. For example, while quantifying the initial atrazine concentrations in treatments is important, the assumed absence of atrazine in controls should not be grounds for automatic dismissal of the study. Knowing the particular decay rate of atrazine may be useful information, but not having this information does not prevent determining the cause and effect relationship between atrazine and the endpoint measured. In the Panel's view, criteria 1-7 should not be considered as absolutely essential elements in judging the quality of the published study for use in risk assessment (i.e., many exceptions were identified), rather the focus should be on comparison of control and experimental treatment(s) and comparisons among treatments. Criteria 8-11 are all contextual in nature (i.e., no definitive "yes" or

"no" answer). The importance of these criteria depends on other factors. The Panel stressed that a study could be considered a high quality study and very useful in risk assessment (even quantitative assessment), even if some of these design elements are not met.

Table A in Appendix C of the White Paper describes the confidence with which the EPA has in the results of each study using the proposed criteria and classification system. The Panel found that the accepted studies marked as "qualitative" were in this category, in large part, because atrazine residues were not measured over time. In addition, some of these studies did not include a description of the solvent used. Other studies in the "qualitative" category used standard animal husbandry protocols that did not meet the EPA's test design criteria (see Panel's response to charge question 4). The Panel provided the following detailed analysis of the test design criteria.

1. EPA criterion: The test should contain both a negative and a solvent control that meet ASTM standards concurrently with the treatment groups.

If a study meets the ASTM standard for use of solvents and if the researcher(s) have shown in previous studies that there is no difference between a negative and solvent control at the concentrations used, then use of one control may be sufficient, but this control should ideally be the solvent. If there is a negative and positive solvent control that exceeds ASTM standards by a small margin, and there's no difference between negative and positive controls at the concentrations used, then the experiment could also be included in the evaluation.

2. EPA criterion: Screening for potential contaminants in food and water sources.

Because thousands of chemicals are purposefully released into the environment, it is unrealistic to screen for a variety of contaminants or interferences in food and water sources that are not being explicitly manipulated. If there is a control and a treatment manipulation, the classic definition of an experiment, this is sufficient to test for effects of atrazine. The current criterion allows for the potential elimination of any study.

3. EPA criterion: Measurement of the test chemical in both the control(s) and the treatment groups to relate potential effects to an exposure concentration and to check for potential contamination in the control(s).

While quantifying initial atrazine concentrations in treatments is important, the assumed absence of atrazine in controls should not be grounds for dismissal of the study. Testing for chemical residue in the control is essential when 1) unanticipated effects are seen in control treatments such as effects on gonads or mortality, or 2) treatment difference(s) are NOT found with controls. Failure to measure the test chemical in the treatment groups makes interpretation of statistically insignificant treatment effects nearly impossible (as it represents a risk of a false negative if the dosing was accidently left out of the exposure), and evaluation of differences among controls and treated groups of limited value for determination of absolute (i.e., quantitative measurement) concentration effects. If significant differences are found in treatments compared to the controls, but test chemicals were not measured in treatments, then determination of the NOAEL (No Observable Adverse Effects Level) or LOAEL (Lowest Observable Adverse Effects Level) will be difficult.

4. EPA criterion: Test equipment should be constructed of recommended materials and avoid the use of materials (i.e., plastics) that may leach contaminants that may interfere with developmental endpoints.

Although testing in glass containers is ideal, it should not be considered as imperative. If there is a control and an experimental treatment(s), comparisons can be made among treatments. This criterion for test equipment material results in elimination of data when control animals appear to exhibit normal responses. We found no available studies that indicate, for instance, that rearing amphibians in plastic results in increased incidence of hermaphroditism or skewed sex ratios compared to amphibians reared in glass. If there is only one difference between the control and experimental treatment, then the difference must be attributed to the manipulated factor.

5. EPA criterion: Loading rates of tadpoles (1 tadpole/L) adheres to recommended rates to ensure adequate water quality and development, in addition to reporting the measured water quality parameters.

The best measure of adequate water quality and development is animals that are healthy and reach normal developmental sizes. Researchers that work with amphibians often rear amphibians together in groups at loading densities greater than 1 tadpole/L; natural densities of amphibians range from 14-4238 tadpoles/1000 L (Morin 1983, Petranka 1989) and can be greater in ephemeral pools or ditches. Amphibian studies have demonstrated that water quality can be maintained with water changes every 3-4 days using multiple animals (Sullivan & Spence 2003). Water changes every 3-4 days in beaker studies, or weekly when using aquaria, is an accepted, common practice among researchers working with amphibians. Odum & Zippel (2008) recommend changing a minimum of 10-20% of the water every one to two weeks to maintain water quality for amphibians. Basic water quality measurements such as ammonia concentrations and pH are recommended to better interpret any observations that might question the health of test organisms.

6. EPA criterion: Plastic test vessels should not be used in the study.

The response observed in the study depends on the degree of concern. If there is a control and a treatment, and a significant difference is observed between treated and control groups, the use of plastic test vessel is less of a concern. Given the potential for some plastics to leach estrogenic compounds and the concern over atrazine's endocrine disrupting effects, it seems plausible that plastic would be more likely to reduce the magnitude of treatment-related differences and, therefore, the likelihood of finding a statistically significant difference between treated and control groups. If there are no studies indicating that rearing amphibians in lab-certified plastic containers results in endocrine disruption for the responses associated with atrazine exposure (the Panel could find none), then the available data should be used. However, if this is a serious concern, these experiments could be conducted by the EPA or independent laboratories to determine if there really is a potential for synergism.

7. EPA criterion: The study is ideally conducted using technical atrazine. Exposure to the additional chemicals may influence the response and confound the ability to discern potential effects from exposure to atrazine alone.

Commercial formulations are applied in nature, and if a formulation of atrazine has a significant effect, then that is relevant. A combination of commercial and technical grade atrazine studies are necessary to evaluate its impacts.

8. EPA criterion: The use of organisms from controlled environments to understand prior exposure history. Laboratory-raised organisms are preferred or organisms from outdoor sources in which information about potential contaminants are known.

Although exposure history could influence a response, it is equally likely that laboratory organisms could be more or less sensitive than animals in nature. Laboratory animals could be more sensitive to stress because of limited genetic variation or because of a limited environmental variation in rearing conditions; conversely, laboratory-raised animals could be more tolerant of stress as a result of selective forces that favor survival in the laboratory. Organisms in nature will have different exposure histories. Ultimately we need to study animals from natural systems to understand the impact of atrazine on animals from a range of circumstances. Such testing would involve species collected from natural and anthropogenic-influenced habitats, as well as laboratory reared organisms, to understand the impact on real populations.

9. EPA criterion: Control mortality should not exceed 30%.

Depending on the experimental venue, mortality could be higher without anything being inherently wrong with the study. In nature, survival of larval amphibians is typically <5% (Semlitsch 1987, Berven 1990, Semlitsch et al. 1996). Therefore, the more natural the conditions (presence of predators, disease, competitors), particularly in experimental mesocosms or experimental wetlands (e.g., Boone et al. 2004), the more likely that mortality could be greater than 30%. Understanding the impact of atrazine in the presence of other factors is essential because biotic and abiotic factors are known to alter toxicity. Given the anticipated high mortality in controls it is essential that this be factored into experimental designs (e.g. appropriate numbers of organisms and replicates).

10. EPA criterion: The presence of other stressors in controls/treatments compromise the interpretation of atrazine as the causal stressor.

Animals in their natural environments are exposed to a variety of stressors and if the goal is to determine the effects of atrazine on a group of organisms (in this case, amphibians) then such studies are valuable for predicting effects in more realistic natural environments. Atrazine can interact with other environmental factors, which can alter toxicity. Studies can be designed to manipulate multiple factors, including atrazine, or to manipulate atrazine exposure under realistic environmental conditions (which may include predators, competitors, pond drying, etc.). Presence of other factors should not be used to eliminate such studies from consideration in a review. If atrazine has an effect, then it demonstrates its impact under those conditions; therefore, the presence of other factors does not compromise the interpretation of the effect of atrazine.

11. EPA criterion: Aberrant effects in controls can confound ability to discern treatment related effects (e.g., high incidence of intersex or skewed sex ratio).

If there are unexpected effects in the controls, it is necessary to confirm that the control is not exposed to atrazine, which may explain the effects. However, there are some studies that suggest amphibians with different life history strategies may in fact show intersex and skewed sex ratios naturally (Storrs & Semlitsch 2008, 2009); therefore, it is possible that intersex or skewed sex ratios could occur in controls for some species. The importance in interpretation of the data then falls on whether the controls differ from the atrazine-treated animals. If so, then the data are valuable and may help identify variability in population responsiveness.

Use of a single *X. laevis* study is insufficient to make conclusions for all amphibians

One of the key issues identified by the Panel is whether the DCI study (two experiments conducted in tandem from the same clutches of eggs from one strain of a single laboratory species, *X. laevis* [Kloas et al. 2009] that met all of the EPA's test design criteria provides a sufficient basis to determine

susceptibility of all amphibians to atrazine. In the White Paper, the EPA made the following statements concerning the use of the DCI study for assessing risk to amphibians: *"Consistent with the EPA's position in the 2007 SAP, the DCI studies provide reliable information on the development, growth and survival of amphibians, using X. laevis as a surrogate, exposed to atrazine. The SAP concurred that atrazine showed no effects on X. laevis at concentrations of 0.01 to 100 ppb. Given the current state of our knowledge, the EPA expects to use these data, i.e., NOAEC=100 µg/L, for assessing risk to amphibians in Registration Review."*

The Panel pointed out that the DCI study may answer the question of whether or not atrazine causes gonadal effects in the strain of *X. laevis* used in the study, but the results of this study are insufficient to make a global conclusion that atrazine has no effect on all amphibian species (this same conclusion was reached by the 2007 SAP). The Panel agreed with the 2007 SAP's statement that the DCI study refutes the modified hypothesis stated in the meeting report (p. 7, SAP 2008), *"the parent compound atrazine causes adverse gonadal development in X. laevis within the exposure concentration range of 0.010 to 100µg/L."* Note: The 2007 Panel did not extrapolate the *X. laevis* results to other amphibians. The EPA's sole reliance on the DCI study to conclude that there are no effects on amphibians at concentrations less than 100 µg/L has a strong potential to underestimate the risk of exposure to atrazine across all amphibians.

What ecological studies have taught us is that species differ in their response to natural and environmental variables in nature, whether it is tolerance to pH, ultraviolet radiation, predators, competitors, temporary environments, or environmental contaminants. Variation in overt toxicity across amphibian species has been demonstrated for pesticide exposure (Jones et al. 2009; Relyea and Jones, 2009). There is no reason to expect less variation in response to the effects of low concentration exposures on endocrine disrupting endpoints. Bridges and Semlitsch (2000, 2001) demonstrated that different anuran species, different populations within a species, and families within the same population of a single species varied greatly in their sensitivity to the insecticide carbaryl. While these results are for carbaryl and not for atrazine, the Panel cited these studies to highlight the importance of including multiple species, multiple populations, and multiple clutches in any assessment evaluating the impact of a contaminant.

With this information in mind, the Panel recommended further amphibian laboratory and cosm studies be conducted on a range of amphibian species, at least three North American species (e.g., a ranid (e.g., Lithobates [=*Rana*] *pipiens*), a hylid (e.g., *Hyla versicolor*), and a bufonid (e.g., *Anaxyrus* [=*Bufo*] *americanus*)) from three populations across the U.S. or the species' range, using multiple clutches per population (at least three), should be tested before a conclusion can be reached that atrazine has no effect on amphibians at concentrations less than 100 µg/L. The Panel suggested that parallel studies be conducted in multiple laboratories using the same test conditions to compare the results of multiple studies to increase the confidence in the conclusions. Note: The 2003 and 2007 SAPs also recommended additional testing on North American species, but such testing had not been initiated prior to this SAP. See also the Panel's response to charge question 4 for a list of species and laboratories that have husbandry protocols that could be adapted for the evaluation of atrazine's impacts on these species.

One Panel member made reference to a current National Toxicology Program study looking at the genetic variability of the rat and mouse models it uses in its toxicology screening panels. This approach is a reflection of our increased understanding that genetic makeup of the animal model used has a major impact on the health effects observed as a response to an exposure. Such recognition of the importance of genetic diversity in evaluating response to a compound again suggests that applying the findings of the DCI study across all amphibian species is premature.

Recommend weight of evidence approach

The EPA performed a qualitative evaluation of the 75 published studies, which indicated a range of effects from no effect to effects on survival, growth, time and size at metamorphosis, terrestrial performance, hormone levels, increased hermaphroditism, sex ratios, and reproduction. The Panel indicated that these studies suggest that atrazine can affect amphibians at low environmental concentrations, although there is variability in the atrazine concentrations that elicited effects among species. While these studies may not be adequate to determine NOAELs or LOAELs, the Panel stated that these data are strong enough to raise concerns for the potential effects of atrazine exposure to amphibians at concentrations less than 100 µg/L. The Panel recommended that a weight of evidence review of the published studies (all 75 published studies), along with additional laboratory and cosm studies on more species (other than *X. laevis*), be conducted before a conclusion is reached on whether or not atrazine affects amphibians at concentrations less than 100 µg/L. Particular attention should be given to atrazine concentrations that impact reproductive endpoints, which appear sensitive to atrazine in some populations and species, and might also impact immune responses, which have not been adequately assessed (see Panel's response to charge question 5). As stated previously, the EPA's sole reliance on the DCI study to conclude that there are no effects on amphibians at concentrations less than 100 µg/L has a strong potential to underestimate the risk of exposure to atrazine across all amphibians. In support of this risk being significant are the data shown in Fig. 10 of the White Paper, where there are more studies that demonstrate effects than do not.

A weight of evidence review, defined as one which uses an accepted set of criteria to determine the utility of each study. A formalized approach to conducting a weight of evidence review is to use meta-analysis, a statistical approach for combining information from different studies to address a research question of interest. Consideration is given to sample size, study design, effect sizes, etc. in the analysis. So, several small studies, which were each under-powered, could show small, but insignificant effects; however, when combined in a meta-analysis, they could reveal that a significant effect occurs. A meta-analysis could use all laboratory and cosm studies that were deemed valid to evaluate the likelihood of potential adverse exposure outcomes and eliminate the bias that may enter into evaluating the results of each study one at a time.

One panel member referred to the recent meta-analysis by Scott et al. (2011) of epidemiologic evidence of trichloroethylene causing cancer, as an example of the use of meta-analysis in risk assessment by the Agency. Another example of a weight-of-evidence approach that differs from the standard meta-analysis approach was presented by Syngenta during the meeting, meeting, although it appeared flawed in its analysis.

The fact that the published peer-reviewed literature is self-correcting over time is important. The Panel noted that there will always be advances in science to improve understanding of the issues. As questions arise from previous studies, and as published papers focus on the remaining questions, there should be a point in time when there are enough answers to determine the range of atrazine concentrations that may or may not cause effects across amphibian species.

b. Please comment on EPA's conclusions about the level of confidence placed on each study's results.

Panel Response

The EPA did a significant review of the literature and captured almost all of the studies that were there with only a couple of exceptions (additional studies are listed in response to charge question 1). As a binning exercise, the EPA reviewers did a fairly consistent job of placing most of the studies into the appropriate categories given the weight the EPA placed on the design criteria. The Panel agreed that most (though not all) papers deemed "invalid" were not of sufficient quality to capture reliable results. Some of the "invalid" studies should be reconsidered as they were removed from the assessment for reasons the Panel stated earlier concerning the strict application of some of the test design elements. Seethe discussion in response to charge question 2a.

Except for the majority of the "invalid papers," the Panel did not agree with the EPA's conclusions about the level of confidence placed on each study's results by retroactively applying the test design elements to evaluate the "acceptability" of a study in the published literature. Of the 75 published studies screened using the proposed test design elements, only the DCI study (two experiments with one strain of a single laboratory species, *X. laevis* [Kloas et al. 2009]), passed all of the test design elements for use in quantitative risk assessment, i.e., NOAEC <100 µg/L. As stated in Panel's response to charge question 2a, a weight of evidence review of the published studies (a meta-analysis), along with additional laboratory and cosm studies on more species, need to be conducted before a conclusion can be reached that atrazine has no effect on amphibians at concentrations less than 100 µg/L (same conclusion as reached by the 2007 SAP).

One implicit expectation that surfaces in the White Paper is that the published studies should show that atrazine produces consistent effects across species at a particular concentration. However, we know (and should expect) that species differ in susceptibility due to genetics (Bridges & Semlitsch 2000, 2001) and/or life history strategies (Storrs & Semlitsch 2008). A number of studies show effects on amphibians from 3 to 30 times below the EPA water standard of 3 ppb (e.g., Hayes et al. 2002, 2003; Storrs-Mendez and Semlitsch 2009) and at levels that species may routinely experience in nature. The available data suggest the potential for endocrine disruption across a number of amphibian species from the lab and the field reared in a variety of conditions, which also appears true for other aquatic vertebrates (Rohr and McCoy 2010, Tillett et al. 2010, Hayes et al. 2011). The Panel was very concerned with these effects and the indication that atrazine is linked to endocrine disruption in amphibians. These data appear to be largely discounted in the current analysis. The Panel recommended that a high priority be given to planning and conducting future studies to evaluate whether atrazine causes endocrine disruption resulting in adverse effects to survival, development, and reproduction (an adverse-outcome pathway). These studies should be conducted in multiple independent laboratories to ensure confidence in the results and their interpretation.

Question 3. After evaluating all the available amphibian studies, one study was found to have accounted for all the identified test design elements (Question #2) and determined to be suitable for quantitative use in risk assessment for the endpoints of survival, growth and development (Section 7.3 and Appendix C). This study was required by an EPA Data Call-In (DCI) Notice following the recommendations from the 2003 SAP on atrazine and amphibians. The resulting study examined the effects of atrazine on *Xenopus laevis* at concentrations of 0.01 to 100 µg/L at two different laboratories. Based on the 2007 SAP, the conclusion was, and there was agreement by the Panel, that the data from this study were robust and sufficient to conclude that exposure to atrazine at concentrations ranging from 0.01 to 100 µg/L had no effect on *X. laevis* development (which included survival, growth, metamorphosis and sexual development).

a. Please comment on whether any new information has become available that leads to a different conclusion from the one which EPA reached in that the results of the DCI study were adequate to evaluate potential effects of atrazine exposure to amphibians.

Panel Response

Aside from Paetow *et al.* 2012, which found a shift in growth rate in *L. pipiens*, the Panel found no other amphibian studies aside from those listed in the White Paper that have tested effects of atrazine on amphibian growth, reproduction and/or survivorship. The Panel disagreed with the EPA's characterization that *"the results of DCI study were adequate to evaluate potential effects of atrazine exposure to amphibians."* Lack of negative effects on *X. laevis* at atrazine concentrations less than 100 µg/L in the DCI study is insufficient evidence to make this same conclusion across all amphibian species and has the potential to underestimate the risk of exposure to atrazine across all amphibians. See the Panel's detailed discussion in response to charge question 2. The 2007 SAP also made similar points (see report, SAP 2008).

The Panel reiterated the following four points made previously in response to charge question 2.

1) The Panel recommended that the EPA's test design criteria be applied in a less absolute fashion in screening the published literature.

2) The Panel recommended that the EPA use a weight of evidence review, i.e., a meta-analysis of all of the existing data. A meta-analysis can remove bias involved in assessing the degree of effects in each study, one at a time.

3) The Panel stated that many relevant studies, where subchronic or chronic exposures resulted in effects on reproductive endpoints, were largely discounted in the evaluation.

4) The Panel recommended additional testing on at least three North American amphibian species to examine whether or not adverse effects occur over a range of amphibians. This recommendation is consistent with similar recommendations made by the 2003 and 2007 SAPs.

b. If such information is now available, please comment on how a threshold determination (a concentration that is expected to cause no effect) may be accomplished using the identified studies.

Panel Response

The Panel examined all of the data from Figs. 7, 8, 9 in the White Paper and found that atrazine caused effects on metamorphosis, growth, and sexual development at 1 ppb and above. These summary results suggest that some amphibian species are more sensitive than primary producers evaluated in Sections B and C of the White Paper. Furthermore, atrazine concentrations below 1 ppb may impact reproduction as Tillitt et al. (2010) demonstrated with fathead minnows. The Panel recommended that determination of the pervasiveness or likelihood of effects at these concentrations and below be considered an important research priority.

The EPA's Fig. 10 (in the White Paper) is helpful in evaluating the overall outcomes of all studies. Although some of the studies in Figure 10 were given lower confidence in the EPA's assessment, this figure suggests two important aspects for broader interpretation: 1) there are significantly more studies

showing effects than not (Fisher Exact P<0.01), and 2) solvent does not appear to influence whether or not an effect was found. As such, the weight of solvent use in evaluating confidence in the studies should be reduced. Therefore, some of the studies that were considered invalid at low solvent levels, e.g., Gunderson et al. (2011), should be re-evaluated to allow for greater confidence in the results. The Panel provided other studies that should be considered in its response to charge question 1.

Question 4. After evaluation of the available amphibian toxicity data, EPA concluded that the DCI study mentioned in Question #3 was appropriate for quantitative use in a risk assessment for survival, growth and development. While the 2007 SAP Panel agreed that atrazine appeared to have no effect on *X. laevis* development at atrazine concentrations ranging from 0.01 to 100 µg/L, they expressed concerns about the suitability of *X. laevis* as a surrogate for native species. Review of the available toxicity data utilizing indigenous species suggests that suitable protocols, including adequate husbandry methods in particular, that would enable EPA to quantify a toxicity endpoint representative of a clear and consistent response from atrazine for native species, may not exist. *Please comment on whether there are suitable methods for testing native amphibians with particular regard to husbandry and laboratory culturing conditions, consistent with the design elements recommended by the 2003 SAP.*

Panel Response

Animal Husbandry

The Panel agreed with the 2007 SAP that extrapolating results from *X. laevis* to North American species may underestimate exposure risk and recommended that the Agency proceed with further studies conducted with at least three North American species (e.g., a ranid (e.g., Lithobates [=*Rana*] *pipiens*), a hylid (e.g., *Hyla versicolor*), and a bufonid (e.g., *Anaxyrus* [=*Bufo*] *americanus*)) from three populations across the U.S. or the species' range, using multiple clutches per population (at least three) should be tested to provide sufficient evidence on whether atrazine causes or does not cause adverse effects to amphibians. This recommendation is consistent with similar recommendations made by the 2003 and 2007 SAPs. As stated in the 2007 SAP report (p. 7), *"…the 2003 SAP noted that the biology of this species (X. laevis) differs in many respects from that of North American species. The 2003 SAP recommended that studies with X. laevis be followed with comparable studies using a North American species as soon as possible. Such comparative studies have not yet been performed."*

The EPA asked the Panel whether there are suitable methods for testing native amphibians with particular regard to husbandry and laboratory culturing conditions and are consistent with the design elements recommended by the 2003 SAP. Although amphibians are not used in routine toxicity testing, and therefore strict standards for culturing and testing have not been developed, several laboratories throughout North America have successfully (and sometimes commercially) cultured native species. . The Panel encouraged the EPA to proceed with testing North American amphibian species and suggested a number of researchers (laboratories) who would be useful resources on how the EPA's test design elements could be adapted to current culturing practices to meet species-specific husbandry needs. Individuals who have experience North American species include: Helbing (University of Victoria; experience with *Lithobates catesbeiana and Pseudacris regilla*), Trudeau (University of Ottawa; experience with *Lithobates sylvaticus and L. pipiens*), Relyea (University of Pittsburgh; experience with Ranidae: *L. pipiens, L. clamitans, L. sylvatica, L. catesbeiana, L. cascadae*; Bufonidae: *Anaxyrus americanus, B. boreas; and Hylidae: Hyla versicolor, Pseudacris crucifer*), and Boone (Miami University, Ohio; Ranidae: *L. pipiens, L. clamitans, L.catesbeiana*; Hylidae: *H. chrysoscelis, H. versicolor, Acris crepitans, P. regilla*).

As discussed in charge question 2, the EPA used test design elements to screen 75 published studies. Several published studies were not considered to be of high quality because current standard practices in amphibian husbandry do not meet some of the EPA's test design elements. Some of these standard practices include: rearing multiple individuals together and water changes every 3-4 days. In general, these conditions resulted in animal performance comparable to field rearing, i.e., high quality conditions. Therefore, the Panel stated that these studies should not be eliminated or placed in a low confidence category because the EPA's test design elements do not appropriately consider the performance standards used in animal husbandry.

The Panel recognized that some of the EPA's proposed test design criteria are inconsistent with current good husbandry practices used by individual laboratories raising native species. The Panel recommended that all test design criteria intended to be used in the animal husbandry community undergo further peer review. The animal husbandry community needs to understand how to design their laboratory studies to meet these criteria. The Panel suggested that this information be published in an appropriate peer-reviewed journal and/or EPA web-sites, and distributed through the amphibian *ecotox list-serv*, as well as disseminated to researchers conducting studies with atrazine and other pesticides.

Mesocosm studies

Mesocosm studies (with 3-5 replicates) are routinely used in amphibian ecology and ecotoxicology, and yield high power, often >0.8, depending on the endpoint of interest and effects size. Typically-used methods are described in Semlitsch & Boone (2009). The Panel agreed with the 2003 SAP's recommendation to conduct field mesocosm studies as effects of atrazine can change between the laboratory and field (SAP 2003). Cosm studies should be conducted under more natural field conditions to examine how natural factors influence the impact of atrazine on amphibians. Such studies will allow for the determination of cause-effect relationships.

The Panel stated that mesocosm studies conducted under more realistic field conditions do not "confound" the ability to discern cause-effect relationships as suggested by the EPA in the White Paper. The Panel acknowledged that increasing the realism of the test environment may result in a loss of control of some extraneous factors, which can reduce the likelihood of finding a treatment effect when there is one. Nevertheless, proper design of the studies can minimize these problems. Information gained from these studies will be quite valuable. The Panel recommended that laboratory or cosm studies should focus on good experimental design, maintaining quality conditions, and incorporating ecological relevance.

Question 5. **A number of studies report the potential for atrazine to modify immune function and infection susceptibility in amphibians (Appendix C). EPA believes the research on these different hypotheses does not provide sufficient data to establish causal linkages among different levels of biological organization to result in adverse effects. Therefore, EPA concluded that a mode of action or adverse outcome pathway leading to effects on amphibian survival, growth or development cannot be established at this time.**

a. Please comment on whether the data in the existing database reasonably supports the hypotheses, or demonstrates that atrazine affects immune function and/or infection susceptibility leading to adverse effects on survival, growth or development; i.e., are there sufficient data to establish an adverse outcome pathway for atrazine effects on

immune function? Please provide a rationale for the Panel's position and discuss the associated strengths and weakness with the data supporting the rationale.

Panel Response

There are relatively few studies to date examining atrazine's effect on immune responses or susceptibility to disease, pathogens, or parasites, making it difficult to determine the importance of these effects. For this reason, the Panel agreed with the EPA's conclusion that existing data are insufficient to support or refute hypotheses that atrazine adversely affects immune functions and infection susceptibility leading to adverse effects on survival, growth or development (i.e., an adverse outcome pathway for atrazine cannot be established).

The Panel considered the available data to be adequate to conclude that there "could" be adverse outcomes and sufficient to justify further investigation into atrazine and its potential impact on disease susceptibility and immune function leading to adverse effects. One path to advancing these discussions would be to convene a more focused discussion within the scientific community to create these hypotheses, identify associated key endpoints that should be measured and criteria for determining what constitutes strong evidence for or against these hypotheses. Research protocols (detailed methods) and especially evaluation criteria resulting from this discussion could be used to "challenge" the scientific community to develop the data needed to focus research priorities. Given the importance disease appears to have in amphibian population declines, this assessment should be a priority.

The Panel provided examples of studies that indicate atrazine can influence disease susceptibility and immune responses by affecting phagocytic activity at 10 ppb (Brodkin et al. 2007), eosinophils at 3 and 30 ppb (Kiesecker 2002), liver melanomacrophage counts at 102 ppb (Rohr et al. 2008), gene expression of immune system genes at 400 ppb (Langerveld et al. 2009), and susceptibility to pathogens (trematodes at 102 ppb (Rohr et al. 2008) and *Ranavirus* at 16 and 160 ppb (Forson & Storfer 2006). The Rohr et al. (2008) study published in *Nature* provides the most compelling evidence to link exposure to atrazine to disease susceptibility and immune responses. Rohr et al. (2008) conducted a field survey examining the effect of 240 factors, including atrazine, on trematode infections. Their analysis showed that atrazine caused a significant increase in trematode infections; additionally, they conducted experimental studies in mesocosms which indicated that atrazine increased trematode infection. Moreover, an analysis by Rohr & McCoy (2010) found that atrazine negatively impacted immune system response and disease/pathogen susceptibility at environmental levels.

b. If the Panel concludes that the existing data are sufficient to formulate hypotheses that atrazine adversely affects immune function and infection susceptibility, but are not sufficient to test the hypotheses (refute or confirm), then please comment on specific study protocols that can be used to test these hypotheses with sufficient rigor to identify effects that can be directly and quantitatively attributed to adverse impacts on amphibian reproduction, growth and/or survival.

Panel Response

The Panel referred to its response to charge question 5a. The available data suggest that atrazine could increase susceptibility of amphibians to disease and affect immune function. Critical endpoints and protocols should be developed by the scientific community (e.g., immunologists). The Panel recommended that research in this area be considered a high priority to address questions concerning potential atrazine adverse effects on immune function and infection susceptibility.

The Panel offered the following recommendations to evaluate potential atrazine effects on disease susceptibility. Although changes in biochemical markers are valuable endpoints, ultimately, demonstrating susceptibility to natural pathogens/parasites is necessary. In particular, amphibians should be exposed to important parasites (trematodes) and disease pathogens, especially the amphibian chytrid fungus, *Batrachochytrium dendrobatidis* [Bd]; *Ambystoma tigrinum* virus and *Ranavirus*, which are affecting species populations and contributing to amphibian population declines (Collins and Crump 2010). The Panel also recommended that the impacts of these pathogens should be evaluated across a range of North American amphibians in laboratory and field mesocosms because atrazine can alter communities and food webs in ways that could make amphibians more susceptible to such pathogens.

Question 6. The cosms were comprised of natural communities of periphyton/phytoplankton; in some cases, vascular plants, invertebrates and vertebrates present in those communities were included in the study (Chapter I, Section 6.1.4). These sources were generally described as streams, lakes, reservoirs, and springs, and are considered to be representative of the structure and function of aquatic plant communities in such water bodies. *Given the diversity of sources and the described communities, please comment on the extent to which these cosm studies taken together provide useful and reasonable physical models of the natural aquatic plant*[3] *communities exposed to atrazine in the U.S.*

Panel Response

The Panel indicated that it is critical to have a common set of criteria to score each cosm study since the calculation of the CE-LOC using the PATI model is highly sensitive to which cosm studies were included in the analysis. This situation does not currently exist and needs to be resolved before a CE-LOC can be determined.

Cosms are useful and reasonable physical models

The Panel stated that "cosms" (a term the EPA uses to describe both microcosms and mesocosms) are the gold standard of evidence by which one can measure the fate and effects of atrazine under a regulatory setting. The 2009 SAP recommended an increased focus on cosm studies since studies in natural streams are difficult to conduct due to the lack of appropriate reference conditions and the confounding factors of physical impacts associated with storm spates and the concomitant conditions of scour/abrasion, high levels of suspended sediments, and increased nutrient concentrations (see report, SAP 2009). Cosms are ideal physical models because they simulate natural systems with the following intrinsic characteristics:

1) cosms can contain multiple trophic levels allowing the measurement of direct and indirect effects of atrazine on aquatic ecosystem structure and function;

2) each trophic level of a cosm can contain multiple species allowing one to measure relative effects across plant species with potential direct comparisons to single species tests conducted in the laboratory; and,

[3] The term 'aquatic plant community' refers to the sum of photoautotrophic (photosynthetic) organisms found in aquatic habitats including microbes such as cyanobacteria and algae present in the water column (phytoplankton) and affixed to surfaces as part of biofilms (periphyton), as well as higher plants such as submerged and emergent macrophyte vegetation.

3) cosms can be used to evaluate factors mitigating the effects of atrazine in natural systems including species adaptation (McGregor et al. 2008), species substitution (DeNoyelles et al., 1982, 1989; Fairchild et al. 1994), functional redundancy (DeNoyelles et al. 1982, 1989; Fairchild et al. 1994), and recovery (Herman et al. 1986; Hughes et al. 1988; Hamilton et al. 1988; Abou-Waly et al. 1991) of plant species and communities (Fairchild et al. 2012). Collectively, these endpoints are considered of high ecological relevance when the goal is to measure the resistance, resilience and recovery of an aquatic ecosystem exposed to atrazine under realistic fate conditions.

Concerns with scoring of cosm studies

The Panel noted that the Agency responded to the 2009 SAP's advice to expand the number of cosm studies used in this analysis by addition of twenty or so new cosm studies. This served to expand the dataset for estimating the LOC. The Panel focused on major concerns with the Agency's scoring process of effect/no effect of the cosm studies listed in Appendix D.

The Panel re-evaluated a subset of the cosm studies in which effects were noted at atrazine concentrations less than 30 µg/L and identified 11 cosm studies with discrepancies that led them to be incorrectly scored as "effects" when they should be really scored as "no effect" (see **Table 1)**. [Note: This is the same finding as in Giddings (2012).] The 2009 SAP flagged a number of the same studies for the same reasons as provided in Table 1. The Panel did not re-evaluate studies where effects were listed at atrazine concentrations greater than 30 µg/L because there was agreement in the scoring.

Five of the studies listed in Table 1 were conducted at the University of Kansas from 1979-1991. These studies were considered "ecotoxicological classics" based on the hypotheses tested, study complexity and ecological relevance. However, they were not conducted under any semblance of Good Laboratory Practices (GLPs). Individual notes for each study are listed in Table 1. Note: The University of Kansas study is accounted for five times in EPA's analysis (i.e., five-fold accounting of the same study with different authorship over a 9-year period), which may bias the data in Fig. 16 of the White Paper for atrazine effects at 20 µg/L.

The Panel recommended that the cosm studies at concentrations less than 30 µg/L (Appendix D, White Paper) be re-evaluated and re-scored using a common set of scoring criteria, e.g., Giddings (2012). The outcome of this entire exercise would be to determine which studies showed effects at concentrations less than 30 µg/L that were truly statistically significant and ecologically relevant. Many of these studies were conducted at the University of Kansas and do not actually contain all of the information necessary to score them since much of the text merely alludes to the original discussion in students' Masters Degree theses or other papers. The Panel would like to see the CE-LOC re-calculated with the re-scored studies to determine the relative effect on the CE-LOC.

Table 1. Summary of the Panel's Evaluation of 11 Cosm Studies

Study Author	Study Evaluation
Lampert et al. (1989)[1]	This study should be excluded due to solvent bias as noted on p. 173 of Giddings et al. 2005; "decreased primary productivity" was actually increased in bacteria growth and respiration (See Appendix D, p. 5). The Panel recommended that this study should be dropped. The Panel was disappointed to see this study still included in the cosm dataset since the 2007 and 2009 SAPs indicated that it be dropped due solvent bias.
DeNoyelles et al. (1982)[1, 2]	This study showed basically no effects on biomass and C-14 uptake in phytoplankton at an atrazine concentration of 20 µg/L. This is revealed by the overlapping confidence intervals in Fig. 1. In fact, the greatest proportional differences occurred when the 20 µg/L concentration simulated primary productivity compared to the control. Figure 2 shows stimulation of 3 species of algae at 20 µg/L. In addition, this study contained gizzard shad at a total of 7 fish/mesocosm, or 70/acre…in addition to bluegill and channel catfish. There was no accounting of survival of gizzard shad which are very difficult to handle in transfer and stocking. Differential survival would have large indirect effects due to differences in the zooplankton community. Effects noted at 1 µg/L were short-term studies where control pond water was treated with atrazine in lab bioassays. No macrophyte data were cited. These data should not be assigned a 20 µg/L effect.
Carney and DeNoyelles (1986) [1]	Authors cited exclosure experiments in control ponds. Grass carp were stocked at 20/acre, which is an order of magnitude greater than common guidance (2/acre). Macrophytes were totally denuded in control pond. Note: These two ponds may not have been the ponds examined by deNoyelles et al. (1982) or Dewey et al. (1985); however, this is the exact magnitude of the direct effect on macrophytes that one would expect at this extreme stocking level.
Dewey et al. (1986)[1, 2]	This study should be excluded because it does not meet the "must not have other stressors present" criterion listed in Appendix D, p. 5. Grass carp, gizzard shad, channel catfish, and bluegill were present with no data on percent survival (especially differential survival of grass carp; however, Kettle et al. (1987) mentioned 80% survival of adult bluegill). Grass carp were stocked at 20/acre, which is an order of magnitude greater that common guidance (2/acre). Author indicated that decreased insect emergence was an indirect effect and not primary effect of atrazine. *Macrophyte biomass was decreased by 90 and >95% in the 20 and 500 µg/L treatments, respectively, which makes no ecotoxicological sense based on the large amount of data for atrazine. Visual observations indicated that periphyton was affected at 100 µg/L but no mention of effects at 20 µg/L.* Note: there was zero grass carp survival in the controls as indicated by DeNoyelles et al. 1989, which resulted in high macrophyte biomass and associated insect emergence rates.
Kettle et al. (1987)[1, 2]	The pertinent reference should be the original 1980 Master's Thesis which led to the 1987 paper. The 1987 paper reports a negative effect of atrazine on bluegill reproductive success. It suffers from the same design flaws as Dewey et al. (1985) such as presence of gizzard shad. With no information on differential survival of gizzard shad the results on bluegill reproduction are suspect. Loss of gizzard shad from the controls would lead to increased numbers of zooplankton and higher survival of young bluegill. Note: there was zero grass carp survival in the controls as indicated by DeNoyelles et al. (1989), which resulted in high macrophyte biomass that served as refugia for young bluegills and allowed them to avoid predation by channel catfish.
DeNoyelles et al. (1989)[1, 2]	Ponds were exposed to 20, 100, 200, and 500 µg/L of atrazine. The results indicate that effects occurred at an atrazine concentration of 20 µg/L. However, the paper explicitly states (Fig. 4) that there were no lasting effects on phytoplankton biomass up to 500 µg/L.

Study Author	Study Evaluation
	Species shifts occurred, but they were replaced by tolerant species. Effects were noted in the laboratory in C-14 uptake experiments, but these are short-time bioassays that do not reflect what happens in the mesocosm itself. Four species of fish were present (bluegill, channel catfish, gizzard shad, and grass carp). Biomass of all species was *increased* in the presence of atrazine with two exceptions: one control pond that had zero gizzard shad survival and another control pond that had zero grass carp survival. Fish data and macrophyte data presented in Fig. 7clearly show the effects of grass carp on macrophyte biomass. Macrophyte biomass in the control ponds was high because grass carp were absent. This observation, not mentioned in Dewey et al. 1986 and Kettle et al. 1987, invalidates these studies as well, as increased insect emergence and bluegill survival were observed in the controls due to decreased predation of young of the year bluegill. In the treated ponds, there was high grass carp survival, macrophyte biomass was decreased due to fish grazing, and emergent insects and larval bluegill were higher due to refugia from predation by bluegills and channel catfish.
Detenbeck et al. (1996)	This study design involved steadily increasing does of atrazine at 2-week intervals from 15 to 25 to 50 to 75 µg/L using two controls and two treatments. The authors concluded that there were effects of atrazine on gross primary production at the 15 µg/L level; however, no data were presented for the initial 2-week 15 µg/L exposure other than two reported dissolved oxygen concentrations. Gross photosynthesis is not reliable in this system due to the accumulation of large amounts of sediment and detritus that would result in high respiration rates. Hence the stressor in the early part of the study was not atrazine, but most likely accumulated decaying organic matter unrelated to the dosing. Neither chlorophyll *a* nor ash-free dry weight of periphyton was affected at any concentration. *Elodea* was not affected at concentrations up to 75 µg/L, but *Ceratophyllum* showed effects only at $\geq$ 75 µg/L. The Panel recommended that this study be excluded from consideration.
Kosinski (1984)[1]	The 10 µg/L atrazine concentration listed for this study in Attachment 3, Appendix D, should be 100 µg/L. Although some streams were colonized (pre-treated at 10 µg/L), there are insufficient data to assign effects at this level. The abstract of Kosinski and Merkle (1984) states succinctly, "There was little evidence that exposure to 0.01 mg/kg herbicide during colonization modified the response of the algae to any of the herbicides."
Sequin et al. (2001a)	Exposure to 2 and 30 µg/L of atrazine had a stimulatory effect on periphyton production, but this was incorrectly categorized as a negative effect.
Seguin et al. (2001b)	No significant effects on periphyton biomass were observed at 30 µg/L of atrazine. The effects observed appear to be based on shifts in phytoplankton community. For example, as atrazine concentrations increased, the number of Chrysophysceae increased while the numbers of chlorophytes decreased. Acetonitrile was used as solvent, but amount added not listed. Rarely is this solvent used in ecotoxicological dosing.
Seguin et al. (2002)	While the results indicate that there was a 30% reduction in algal biomass over a 21-day exposure to 30 µg/L atrazine, recovery was not studied. The preponderance of evidence in the literature indicates that recovery would be expected.

[1]Studies discussed by the 2009 SAP as having flawed methodology, which affected interpretation of the results.

[2] Studies conducted at the University of Kansas from 1979-1991. These studies were considered "ecotoxicological classics" based on the hypotheses tested, study complexity and ecological relevance. However, they were not conducted under any semblance of Good Laboratory Practices (GLPs). Note: The University of Kansas study is accounted for five times in EPA's analysis, which may bias the data in Fig. 16 of the White Paper for atrazine effects at 20 µg/L.

Issues to be considered in the evaluation of cosm studies

A key consideration in evaluating the suitability of the cosm studies and the single-species toxicity studies used in developing the CE-LOC is whether they represent the full range of primary producer sensitivity to atrazine. The single-species studies were based on taxa that are straightforward to culture in the laboratory, but which also tend to be found in more anthropogenically "disturbed" sites. As such, these species would be considered generally "tolerant" (that is, tolerant to various forms of anthropogenic stress, which could, but does not necessarily, include atrazine). Furthermore, based on the stated data-selection criteria for the problem formulation, one cannot be certain that none of the primary producer species used in the toxicity tests (or in the mesocosms) were atrazine-resistant. Use of tolerant (or resistant) species runs the risk of underestimating negative effects of atrazine, if one makes the reasonable assumption that they occupy the range of taxa that are less likely to respond to the herbicide. Similarly, the primary producer assemblages used in cosm studies, if sourced from sites that have a history of anthropogenic impacts (which may or may not include atrazine), might not exhibit the full range of effects expected with more sensitive assemblages that would be found only in relatively unimpacted "reference sites" (described below). These factors taken together could result in underestimating the effects of atrazine in natural primary producer communities.

With the noted exceptions above in responses to charge questions 2-4, the plant community is expected to be the biotic assemblage most sensitive to atrazine, such that setting the CE-LOC based on response of the plant community will by extension be protective of the ecosystem at large. The White Paper states, *"By focusing on aquatic plant community structural and productivity changes, EPA intended to protect invertebrates, fish, and amphibians from the direct effects of atrazine as well as the effects that atrazine could have on the habitat and food sources of aquatic animals"* (p. 8), *and, "While the LOC is based on effects to aquatic plant communities by ensuring protection of primary producers, it is intended to provide protection for the entire aquatic ecosystem including fish, invertebrates, and amphibians"* (p. 43). The Panel expressed that it was unclear whether the use of only a generally tolerant group of primary producer species would achieve these ambitious goals.

If the goal of the risk assessment is solely to determine whether atrazine will have (further) detrimental effects on primary producer communities that already have been exposed to anthropogenic stress (which may or may not include atrazine), then perhaps it is sufficient to use generally tolerant species for the toxicity tests, and it is likewise acceptable to source material for the cosm studies from sites that are subjected to human disturbance of various forms. However, if that is the case, then the meaning of the CE-LOC and what it purports to protect should be qualified accordingly. For instance, it should be made clear to all concerned that the CE-LOC is not established as being protective of primary producer communities in relatively pristine sites, because such an assertion will not be supported by the manner in which the CE-LOC was developed.

If, on the other hand, the goal of the CE-LOC is to represent effects on primary producer communities in a broad sense, then a broader representation of primary producer species should be employed for its determination. The White Paper (p.91) states that *"The LOC methodology uses single-species plant toxicity data and microcosm/mesocosm (cosm) studies to determine what atrazine exposure patterns and concentrations can cause adverse effects on aquatic plant communities,"* suggesting that this is the case. "Reference sites" are often used in studies that seek to determine the full range of potential effects of one or more stressors on ecological response variables. In the present case, the stressor is atrazine, and the response is in terms of primary producer (community) endpoints. Within the context of the risk assessment for atrazine, the reference concept could be used as a guide for identifying "sensitive" species to include in toxicity studies and locations for sourcing primary producer material to use in cosm studies and/or locations at which to conduct field studies. Ideally, reference sites should be selected

from within a similar geographic, topographic, climatic, geological, etc. setting to the region where the effects under investigation (i.e., atrazine use) occur, such that species native to the region, and therefore most relevant to the study, are used. Identifying such sites (e.g., finding relatively undisturbed sites within atrazine-use areas) is not necessarily an easy task. However, various options (with associated caveats) exist, such as locating "best-available," not pristine, sites (see US EPA 2005 and Stoddard et al. 2008). At least some members of the panel believe that explicit inclusion of reference sites would result in a better representation of primary producer sensitivity to atrazine and a more defensible CE-LOC. This will be possible if new studies are conducted, as per the suggestion of the Panel.

The Panel identified a number of specific "issues to be considered" when evaluating cosm studies. During the course of the meeting, the EPA asked the Panel several clarifying questions regarding the use of the cosm studies. Responses to these questions are included in the discussion below.

1) **Alternative filtering criteria.** The Agency requested clarification on how to evaluate the existing cosm studies. The Panel offered the following filtering criteria for evaluating cosm studies.

   Criterion 1: Lotic or lentic systems. Use only cosm data from lotic systems (e.g., artificial streams, flow through systems) for use in evaluating responses in lotic systems.

   From the AEEMP data sets, conduct a statistical analysis that determines the mean duration of peak atrazine concentrations in vulnerable watersheds, the average time between multiple peaks, and the overall period during which atrazine in stream water is elevated. Make a reasonable decision on what maximum duration of a cosm study is a reasonable model of atrazine exposure to natural communities in vulnerable watersheds. Accordingly, from the same dataset one could make a statistically meaningful determination of the maximum concentrations of atrazine that can be expected to be encountered and use this concentration to exclude cosms that have atrazine concentrations that exceed this by some reasonable factor (e.g., 1.5).

   The Panel noted that all cosm studies have some meaningful information to provide. Prudence should, however, be exercised to select cosm studies that are deemed reasonable models of atrazine exposure to natural communities in vulnerable watersheds based on these additional selection (filtering) criteria.

2) **Time of study duration and realistic exposure conditions.** The Panel expressed concern over the issue of time of study duration. For example, in Figure 16 of the White Paper, approximately 10% of the cosm studies were conducted for 100 days or longer. This time frame is considered unduly long given the patterns of atrazine contamination evident from abundant chemographs available from the monitoring data sets, the scale of growth responses by photoautotrophs, and the evidence for recovery from photosynthetic inhibition by atrazine upon relaxation of exposure. Additionally, even though one cosm study had a duration of 365 days, the actual endpoints were assessed only during the growing season, i.e., 90-120 days. The Panel recommended appropriate care should be given to assigning accurate exposure duration (assessment period) to each study in the cosm dataset.

3) **Realistic atrazine concentrations.** The Panel noted that 10% of the cosm studies have atrazine concentrations that far exceed observed peaks of atrazine in vulnerable watershed streams – even those due to atrazine spills. These should be discounted on the basis of unrealistic conditions

that do not reflect reasonable models of atrazine exposure to natural communities in the United States.

4) **Under-represented lotic studies.** The Panel indicated that attention should be paid to under-represented studies (or lack of studies) in the current database. For example, lotic studies that simulate 2nd-3rd order streams appeared lacking according to the cosm study list in Attachment 3 of Appendix D. Only 30% of the cosms listed were lotic (running water). No studies have been conducted that simulate non-perennial streams, e.g., Nebraska AEEMP sites that intermittently go dry. Such studies would address the questions: "What is the added effect of desiccation on stream communities in addition to the effect of atrazine? Does desiccation 'reset' the community separately from atrazine, thereby facilitating recovery due to intrinsic processes in headwater streams?" Additional studies in lotic systems to address these questions are recommended.

5) **Bias in choice of plant species.** The Panel expressed concern regarding the choice of photoautotrophic species used in laboratory and cosm studies. Most studies used species that were easy to culture or obtain from commercial sources which may bias studies to relatively resistant species. The Panel suggested that the EPA consider whether other species not in the database should be added to determine the accuracy of effect estimates.

   Splitting versus lumping of effects within a single study. Another clarification requested by the EPA was in regard to the concept of splitting vs. lumping of effects within a study (e.g., populating the cosm data set with different taxa and concentrations within a single study). In response to this question, the Panel pointed to the description on slide #55 in Dr. Giddings's presentation on how certain endpoints in the cosm database could be split, deleted, rescored, or added in order to better populate the cosm database, noted as alternate cosm sets 1to 5 (inclusive) (presentation is in the public docket, EPA-HQ-OPP-2012-0230). Graphical representations of the alternative datasets were provided on slide #89. Splitting of the data reduced the number of Effects Exceedance Factor (EEF) scores greater than 1 by 45% compared to the EPA's 2012 dataset (scaled at 100%). A similar reduction was observed with the SAP 2009 recommended dataset (deletions and additions), a range of 40 to 60% reduction in exceedances. Moreover, the simply rescoring approach reduced the number of exceedances by 75%. The fact is that any of the alternative datasets chosen would reduce the number of exceedances remarkably.

6) **Recovery.** The Panel noted the critical importance that less than 10% of the cosm studies listed in Appendix D demonstrated "recovery," and most studies were not designed or conducted long enough to measure recovery of the photoautotrophic community. Although the Agency's presentation indicated that recovery after atrazine exposure is defined as "*a return to pre-exposure conditions for the affected individual, population or community, not for a replacement population or community of more tolerant species,*" the Panel indicated that this definition lacked realism since photoautotrophic communities are dynamic and always changing throughout the season due to competition for nutrients and changes in temperature and light regimes.

7) **Changes in diversity.** The Agency requested clarification on how to determine the ecological significance of "changes in diversity." Changes in biomass are the primary endpoint by which significant changes are determined to have occurred in response to atrazine exposure. Rare, spurious significant changes in biomass would not necessarily be determined to be ecologically relevant since they can and do occur by chance. Relevance is determined as to whether or not a statistically significant effect is observed consistently over the course of the exposure. In practice, however, consistency is dependent on how often biomass is measured. One suggestion

was to define a level of concern for change in diversity, where changes are observed at 25% of observations with special significance given at the end of the exposure, e.g., statistically significant 25% biomass reduction. However, even this may not be detected without adequate statistical power of the test.

Recovery of biomass will depend on the taxonomic group of photoautotrophs that are being considered. Planktonic algae and cyanobacteria should be expected to recover within 10-14 days post-exposure; periphyton recovery should occur within 14-28 days; and macrophyte recovery within 1-2 months during the months of May, June, and July. Total recovery of macrophytes may not occur if the exposure occurs in August or September, since at this point macrophytes begin to senesce due to periphyton shading and competition. Macrophyte recovery times should be independent of exposure duration since plants are known to quickly return to normal function after atrazine stress is removed but recovery of biomass may take longer than phytoplankton and periphyton.

The ecological significance of changes in biodiversity is admittedly more difficult to judge. There are currently no commonly accepted models or criteria to define an "unacceptable change" in biodiversity. There are, however, practical examples of changes that one would consider as ecologically significant such as the loss of a critical macrophyte species important as habitat to amphibians or in maintaining the stability of a stream bank to erosion. Changes in periphyton species composition may or may not be ecologically significant. Periphyton is an assemblage of cyanobacteria, algae, diatoms, and filamentous algae with associated heterotrophic bacteria, fungi, protozoa, rotifer, and invertebrates that can take many different physical forms. It is generally accepted that diversity of invertebrates, like many other species, will increase with habitat complexity. If the periphyton community is changed to a different three-dimensional physical structure (e.g., erect and vertical versus flattened, horizontal forms) then taxonomic changes may result in alteration of habitat for micro and macro-invertebrates. This would be ecologically significant. However, these types of changes in three-dimensional forms are the same responses that could result from storm events due to scouring and abrasion or grazing effects.

Hardest to judge is the ecological significance of a change in species composition of periphyton when the three-dimensional structure of the community is not changed. As long as the three-dimensional matrix is retained, then micro-invertebrate habitat is not changed. There was some disagreement among panel members in terms of potential significance of change in species richness or species composition. Several panelists stated that the significance of such changes would be more statistically relevant than ecologically relevant, e.g., substitution of resistant diatoms for sensitive algae in a periphyton community would not be considered ecologically significant as long as biomass and productivity are maintained. Other panelists stressed that other relevant factors could still come into play, such as whether the newly colonizing species are as nutritious or ingestible as those they replaced. If not, then negative effects on higher trophic levels could ensue, resulting in diminished ecosystem integrity.

The Panel further noted that the AEEMP stream sites were likely already composed of "tolerant" species compared to reference conditions. Although "tolerance values" are classically assigned in relation to nutrient sensitivity and not atrazine, this is significant because if the streams studied were already exposed to high levels of anthropogenic stress in general, it is not clear that relatively small differences in measured atrazine concentrations among the 3 streams highlighted in the pilot field study would be expected to exhibit detectable effects attributable solely to

atrazine. Any potential shift in species composition will likely be from one tolerant species to another within the same phylogenetic group (e.g., Chlorophyta, Chrysophyta, Cyanophyta, etc.). Numerous studies have shown that such shifts in most cases occur with minimal change in system function (e.g., primary productivity, nutrient cycling, and physical habitat quality). Thus, the importance of establishing a valid reference condition is of paramount importance to proper evaluation of cosm studies.

One panel member suggested that assessment of cosm studies move away from a binary decision to more of a "grading of the strength of evidence" decision. As the EPA begins to explore the use of concepts such as "ecosystem health" and "ecosystem integrity," it will be difficult to make the binary decision as to whether "health" or "integrity" have definitely been changed or impacted. Using a grading system will help make the decision process easier and allow experts on peer-review panels to express their level of certainty better. While, we may be able to decide with certainty whether atrazine impacts a particular member of the exposed plant community, whether loss of that member adversely impacts ecosystem health or integrity is much less certain. This explains why there is a need for a "grading" system.

Finally, the Panel commented that the evaluation of the ecological significance of a change in diversity will vary among researchers depending on their background. Ten ecologists would likely give you 10 different opinions on a broad continuum of response. This is why the Panel indicated that true consensus can only be obtained with input from a much larger group of experts in a 2-3 day meeting. This may be the best approach to develop qualitative and quantitative triggers of what might be considered significant ecological effects of changes in species diversity.

Defining the characteristics of a high quality, quantitative cosm study

The Panel indicated that the design elements required for a high quality, quantitative cosm study provided by the EPA in the White Paper (see Appendix D, Attachment 1, p. 5-8) were quite comprehensive and serve as the basis for designing future studies (see section 6.5, *"Specific Panel Recommendations and Paths Forward"*). The Panel provided the following set of general characteristics that would help define high quality, quantitative cosm studies.

1) **Natural conditions.** Cosm studies should be designed to more precisely reflect natural field conditions and examine how spatial scale, time, water flow, etc. interact to influence atrazine toxicity. Furthermore, they should be conducted using a rigorous experimental design with a minimum of three replicates for each treatment to more explicitly determine which natural factors influence the impact of atrazine.

   Since determining effects on animals in the field is the ultimate goal, so more realistic studies need to be conducted, because environmental conditions could influence the impact of atrazine—and comparing animals exposed to atrazine with controls under more natural conditions will allow for the determination of cause-effect relationships. Mesocosm studies are routinely used in amphibian ecology and ecotoxicology and yield high power (often ~0.8, depending on the endpoint and effects size, with 3-5 replicates); typically-used methods are described in Semlitsch & Boone (2009).

   The 2003 SAP recommended conducting field studies, as effects of atrazine can change between the laboratory and field (i.e., field mesocosms). The presence of natural factors increases the realism of the study and does not "confound" the ability to discern cause-effect relationships as

suggested in the white paper. The Panel was aware that increasing the realism of the test environment (e.g., mesocosms) can also result in a loss of control, which will increase the opportunity for effects from extraneous factors to impact results and increase the uncertainty, unless higher sample sizes are used. Typically costs per experimental unit increase and the inclination is to have fewer samples rather than more.

2) **Cosm size.** The size of the cosms should be related to the particular aquatic plant community of interest. For example, cosm studies focused on phytoplankton can be as small as 15 liters, where the community can be sampled several times a week for biomass and species composition over a 14- to 21-day period without study bias due to the small amount of water removed for measurements. In contrast, cosms used to study macrophytes should be greater than 1 m^3 to eliminate problems with nutrient depletion and self-shading which can bias results. Finally, cosms incorporating fish should be of a sufficient size as to allow robust growth of fish without excessive intra-specific competition that will substantially alter plant communities (e.g., 2 gm biomass/m^3; see Touart 1988).

3) **Design reflects type of water body.** Cosm designs and hydrologic regimes should reflect those which are characteristic of the water body type(s) of interest.

4) **Cosm duration.** The duration of a quality cosm study should match the simulated chemograph of interest as well as the life history (e.g., life cycle with allotted time for recovery) of the plant groups relevant to the watershed of interest. Consultation of chemographs from existing monitoring data would be particularly useful for devising exposure regimes for use in cosms, in order to simulate realistic, and therefore more relevant, exposure regimes.

5) **Broad representation of plants.** A broader representation of plants, representing the range of potential levels of sensitivity to anthropogenic stress (including to atrazine) should be employed for determination of the CE-LOC.

6) **Source of re-colonization.** Quality cosm studies should incorporate a source of re-colonization. This source of re-colonization can be internal (e.g., natural sediments and associated microbial systems) or external (e.g., sequential addition of algae, periphyton, or macrophytes). Natural sediments, with associated bacterial, algal, and periphyton communities are essential because they allow the cosm to develop critical components of structure (e.g., species richness, diversity, biomass, and productivity) and function (e.g., productivity, system metabolism, nutrient cycling). Collectively, these attributes make the cosms excellent physical and biological models to assess the fate, effects, and recovery of aquatic systems exposed to atrazine.

Specific recommendations and paths forward

The Panel provided the following conclusions and recommendations to advance the analysis of existing cosm studies and design of new studies.

1) **Re-evaluate and re-score existing cosm studies with purported effects identified at atrazine concentration less than 30 µg/L.** Most of the existing cosm studies with purported effects identified at atrazine concentrations less than 30 µg/L have weaknesses in their design that render interpreting their results and scoring them for "effects" or "no effect" difficult and subjective. The Panel recommended that the cosm studies at concentrations less than 30 µg/L (Appendix D, White Paper) be re-evaluated and re-scored using a common set of scoring criteria, e.g., Giddings (2012). The Panel's re-evaluation of this subset of cosm studies identified 11

cosm studies mis-scored by the EPA as having "effects," when they should have been re-scored as having "no effect" (i.e., less than 30 µg/L, see Table 1). The Panel suggested that these studies be re-scored and/or dropped from the dataset. The 2009 SAP made a similar recommendation. The Panel noted that EPA would not have to re-evaluate this subset of cosm studies if they adopted the scoring of Giddings (2012). This would be appropriate since the rescoring by the Panel of this subset of studies agreed with that of Giddings (2012). The outcome of this entire exercise would be to determine which studies showed effects at atrazine concentrations less than less than 30 µg/L that were truly statistically significant and ecologically relevant. Many of these studies were conducted at the University of Kansas and do not actually contain all of the information necessary to score them since much of the text merely alludes to the original discussion in students' Masters Degree theses or other papers. Studies at atrazine concentrations greater than or equal to 30 µg/L do not need to be re-scored because there is agreement on the scores. The Panel would like to see the CE-LOC re-calculated with the rescored studies to determine the relative effect on the CE-LOC.

2) **Additional focused cosm studies are needed to specifically address the durations of exposure expected at environmentally relevant concentrations.** The Panel overwhelmingly agreed and strongly recommended that additional focused cosm studies are warranted to address the durations of exposure expected at environmentally relevant concentrations. Results from these studies would be used to fill in the data gaps that currently exist for determining whether there are statistically significant and ecologically relevant atrazine effects at the 4 to 7 µg/L level (proposed as the CE-LOC in the Agency's White Paper) and also at higher concentration levels.

   The Panel suggested that a cosm study be developed that meets the criteria for a high quality, quantitative study (described above). The design of the cosm study should be based on the exposure scenario for a 2^{nd}-3^{rd} order stream (i.e., lotic system) with an emphasis on periphyton, and using a set of primary producer species representing a broad range of sensitivities to anthropogenic stress. The same set of measurement endpoints used in the Syngenta field reconnaissance studies conducted by Andrus et al. in Missouri and Illinois presented to the Panel would be a good start in the design of the study(see Andrus et al. presentation in the public docket, EPA-HQ-OPP-2012-0230). In addition, the study should be conducted using a replicated experimental design with concentrations that bracket the 4 to 7 µg/L proposed CE-LOC (i.e., 0, 4, 8, 16 µg/L) and extend to concentrations which are more likely to have statistically-significant ecological effects (i.e., 32, 64, and 128 µg/L). Currently, there are no lotic studies at this range of concentrations in the database.

3) **Proceed with evaluation of atrazine as the cosm data are refined and improved.** While recommending further cosm studies, the Panel also stated that the Agency should proceed with the evaluation of atrazine as plans are developed for implementing further studies. Specifically, the re-evaluation of cosm studies may lead to a less or greater need for the new studies than presently perceived.

Question 7. The Aquatic Plant Community CE-LOC methodology for atrazine is a *four* stage process that uses single-species plant toxicity data and cosm studies to discern atrazine concentrations and exposure durations that may cause adverse effects on aquatic plant communities. As a result, a CE-LOC for atrazine is developed which, together with monitoring data, can be used to identify watersheds where concentrations may result in adverse effects to aquatic plant community structure, function, and/or productivity. *Please comment on the methodology EPA has used to derive the atrazine CE-LOC for aquatic plant communities, and in particular on EPA's characterization of the uncertainties and assumptions in this methodology (Chapter IV, Sections 13 & 14).*

Panel Response

The Panel was very impressed with the amount of work and consideration put into these calculations and estimates. The Panel agreed that the four stages of the aquatic plant community LOC methodology for atrazine were logical. Nevertheless, the Panel expressed genuine concerns with aspects of the LOC methodology, notably the structure and assumptions of PATI and scoring the cosm data. As a result, the Panel expressed minimal confidence in the calculated CE-LOC.

In brief, the LOC methodology has four stages summarized in Figures 15, 16 and 17 of the White Paper. The first step is an evaluation of the single-species plant toxicity data and cosm studies using a binary classification system of effects versus no-effect (Brock scores of 1 and 2 = 0 = no effect; Brock scores 3, 4, and 5 = 1, an effect). Following this analysis, a Plant Assemblage Toxicity Index (PATI) is developed – analogous to a type of Species Sensitivity Distribution (SSD). PATI is an average toxicity relationship across an assemblage of plants. The PATI scores, calculated as the mean % reduction in growth rate across the assemblage, are iterated to establish daily values and cumulated over a 60 day period (a cumulative PATI score). The PATI-LOC is established using the cumulative PATI scores. Later in the process, watersheds that may exceed the PATI-LOCI would be identified from chemographs representing real-world exposure scenarios (monitoring data).

Structure and assumptions of PATI

The 2007 and 2009 SAPs discussed at length the challenges associated with an extension of the cosm data to the real-world (see reports, SAP 2008, 2009). The major challenge is to use the PATI-derived LOC, as calibrated via defined cosm exposure profiles to identify watersheds with complex environmental interactions and variable atrazine exposure profiles that may pose a risk to the structure and function of the aquatic plant community. The Panel did not think this fundamental challenge had been adequately addressed in the present application of PATI. As pointed out by the 2009 SAP, PATI does not account for interactions within and between species (SAP 2008). Thus, for example, if a 100 ppb short duration exposure wipes out a species and a new species takes over that niche, will the system ever regain its original structure?

The EPA White Paper describes the key assumptions used in the cumulative PATI to classify toxicity of atrazine exposures. As stated in the White Paper (p. 118), *"....the community level toxic effect of atrazine can be characterized by the relative reduction in the average specific growth rate for the plant assemblage."* The Panel discussed the structure and assumptions of PATI in some detail.

The Panel noted the importance of understanding that for a given atrazine exposure, the PATI estimated average Specific Growth Rate (SGR) reduction does not correspond to the reduction in the fractional rate of change of biomass of the aquatic community with time. Estimating the "SGR" of the community would require, in addition to knowing the toxic effect of atrazine on individual species, knowledge of

the relative mass fraction of each species. The PATI daily value represents an average, relative reduction in SGRs assigning an equal weight to all species in the assemblage. Furthermore, the Panel could think of no means to actually measure the certainty in the model's predicted adverse effects, and thus, there were no means of ensuring the model provides reasonable estimates of toxic effects. Without this objective basis to evaluate the model, the model loses its credibility (see additional discussion on "objectivity of PATI" below).

The Agency performed an uncertainty analysis to characterize the influence on PATI estimates due to variability in atrazine toxicity measurements for single species, among four major taxa (e.g., vascular plants, diatoms) and due to various taxonomic weightings, etc., and concluded that, despite the large sensitivity of PATI to these uncertainties, the impact on the LOC was negligible because the latter is primarily based on the cosm studies. As stated on page 29 of Appendix E, *"PATI is only being used to assess the relative effects between different exposure times series, and these relative effects are similar whether the plant assemblage is sensitive or tolerant...these relative effects are related to…slopes…which differ little among the various taxonomic assemblage definitions compared to the large variation in the absolute PATI values. However, there are still some effects of taxonomy on EEFs because PATI is not linear with concentration."*

The Panel discussed the relationship between PATI (%) and atrazine concentration illustrated in Figure 17 in Appendix E (p.31), which provides a plot of atrazine concentration on a logarithmic scale (abscissa) versus PATI (%) on a linear scale (ordinate). This figure shows that the relationship between PATI (%) and atrazine concentration is not linear on a log-linear scale. The Panel indicated the relationship between atrazine concentration and PATI (%) across all concentrations will also be non-linear when viewed on a linear scale for atrazine (figure not shown). However, if one plots the untransformed environmentally relevant concentration range for the data, i.e., atrazine concentrations < 100 ppb, and the PATI (%), the resulting relationship is statistically indistinguishable from a linear relationship. (Note: This linear relationship does not hold across all values of atrazine concentration, e.g., atrazine concentrations > 150 ppb.) To illustrate the linear relationship, the Panel transformed the log-linear plot (Fig. 17) to a linear plot by overlaying a dense grid on the physically relevant section of Fig. 17, atrazine concentrations < 100 ppb, and read the PATI (%) values corresponding to 2, 5, 10, 20 50 and 100 µg/L. The resulting pairs are given in **Table 2** and the grid overlay is shown in **Figure 1.**

Table 2. Pairs of Estimates of PATI (%) and Concentration (µg/L)

PATI (%)	0.9	2.2	4.4	10.0	22.2	43.0
Concentration (µg/L)	2	5	10	20	50	100

.

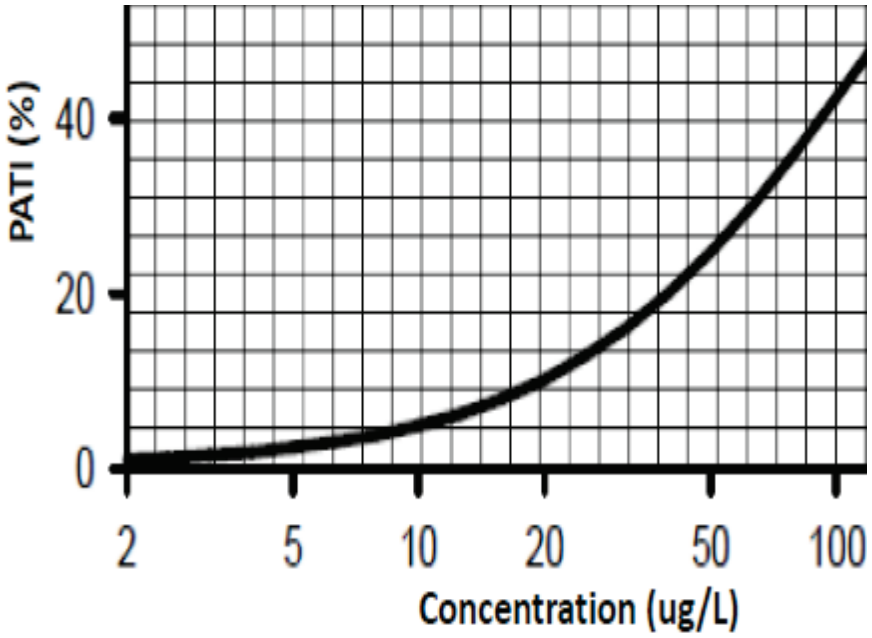

Figure 1. Grid overlay

The following linear regression equation was estimated from these data, which shows that the PATI (%) versus atrazine concentration relationship (for atrazine concentrations ≤ 100 ppb) is extremely well characterized as a linear relationship.

$$PATI(\%) = 0.43 * Conc_{Atr} + 0.4$$

$$Coefficient\ of\ Determination = r^2 = 0.99$$

An additional assumption of PATI is that the estimated toxic effect is equivalent whether the cumulative average specific growth rate reduction occurs within one day or over multiple days.

Consider these assumptions with respect to two hypothetical 60 day chemographs. In chemograph 1, an atrazine concentration of 100 ppb was observed over a 3 day period, with non-detects thereafter. This exposure profile resulted in a ~43% daily reduction in the average specific growth rate (i.e., PATI score) during the first three days and a cumulative PATI score for the 60 day period of 130% (approximately equal to the LOC_{PATI} of 132%). In chemograph 2, an atrazine concentration of 4.2 ppb was observed every day over the entire 60 day period, which resulted in a relative daily reduction of 2.2% in the average SGR and a cumulative PATI score of 132 %. The PATI model assigns the same PATI score to both chemographs, which indicates that they have equivalent adverse, irreversible effects. If similar assumptions were made with regard to alcohol consumption, for example, the implication would be analogous to assuming that the toxic effect of drinking six beers in one hour and then abstaining for the subsequent 59.9 days is equivalent to the toxic effect of drinking one beer every 10 days for the same two month period (with the caveat that the adverse effects on the community of beer drinkers could conceivably be measured). The PATI model assumes the adverse effects are equal for all 60 day chemographs which have the same dosage value. Dosage is defined in the following equation with the assumption that there is a linear relation between PATI (%) and atrazine concentration which is representative up to a maximum daily average concentration ≤ 100 ppb.

$$dosage(t_{60\,d}) = \int_0^{t_{60\,d}} C_{atr}(\tau)d\tau$$

The Panel expressed concerns with the implication of these assumptions. First, there is no clear basis for validating the model predictions, and second, there is no clear basis for the resulting assumption that the adverse effects on a community of aquatic plants follows this linear dose-response relationship (i.e., it is not clear that the adverse effect depends only on dosage per se and is essentially independent of the time course of the exposure). If the community level dose-response is non-linear, then the PATI-based LOC may overestimate adverse effects due to relatively uniform chemographs and underestimate adverse effects associated with highly variable chemographs.

The EPA estimated the probability of an adverse effect occurring due to atrazine exposure profiles in the cosm studies. This probability was derived using logistic regression to examine the significance of changes in various endpoints including aquatic community diversity, productivity, structure and function. However, these studies do not provide a measure of the relative change in specific endpoints as a function of exposure profile, rather they provide a probabilistic measure of observing a composite endpoint adverse effect (as defined above some minimal effects level). The EPA calibrated PATI to this probability measure (e.g., P(0.5) of observing an adverse effect) and then used PATI to estimate the relative reduction in the average SGR of a defined plant assemblage due to differences in atrazine exposures (i.e., variable chemographs, with dynamic flow and environmental conditions). The Panel indicated that there were no observational data to evaluate the validity of the resulting predictions, and thus, determine if this extrapolation is reasonable.

Perhaps this deficiency can be overcome if weighting specific species or taxa is used in the modeling. For example, Andrus et al. (presentation found in the public docket, EPA-HQ-OPP-2012-0230) presented several field bioassays that attempt to study the impact of exposure profiles on aquatic communities. This path may be an objective instrument to evaluate a modified version of PATI which assigns relative weights to different species. There would be questions concerning the choice of controls, randomization and replication with this approach.

The Panel discussed the use of the PATI approach proposed by the EPA versus the use of ppb-days proposed by Syngenta. Both approaches make the assumption that the adverse effects experienced by an aquatic community due to atrazine exposure on any given day are independent of prior exposure history. The Panel had several concerns with this assumption relative to the suite of adverse effects endpoints of interest. For example, the potential for an adverse effect should increase if prior exposure history has altered community structure or increased availability of carbon, nutrient availability, or decreased oxygen availability, etc., despite the apparent rapid reversal of atrazine impact on photosynthesis. In response to a presentation made during the public comment period by Dr. Brain, Syngenta (presentation found in the public docket, EPA-HQ-OPP-2012-0230), the EPA noted that the SGR following atrazine exposure did not reach the initial control SGR within two days after cessation of exposure. In such cases, the Panel noted that the relative change in observed adverse effects in response to the additional exposure appeared to be significantly different than if there had been no prior exposure history.

The Panel also noted that SSDs derived for lab studies would differ from SSDs expected in streams due to differences in environmental factors, e.g., physicochemical variability, light, etc.. With this in mind, PATI, which relies on laboratory based SSDs and SGRs, may differ significantly from SSDs and SGRs relevant to 2nd and 3rd order streams. In Appendix E (p. 7) of the EPA White Paper it states that *"The*

(PATI) methodology is not intended to address…factors, such as physicochemical conditions and the nature of the biological community. Addressing such conditions is not feasible from a standpoint of both effort/cost and knowledge of their influence on atrazine effects." However, such factors have been readily shown to have a significant impact on adverse effects, e.g., the adverse effects of atrazine exposure are greatly reduced as light intensity decreases (Brain et al. 2012). The Panel posed the following two questions, "How can the PATI approach be used to estimate the impact of physicochemical variability on predictions of adverse effects?" and "Could CASM-atrazine provide a relative measure of the impact of such variability on atrazine toxicity?" In addition, the analysis need not be site-specific, but rather along the lines of what the Agency did in section 4 of Appendix E, in which the impact of uncertainty in toxicology, taxa, etc., is evaluated. The Panel recommended that the variability due to interactions of environmental factors with SGR response to atrazine exposures be assessed. As the 2009 SAP noted, there is a distinction between spatial and temporal variability and uncertainty (see report, SAP 2009). In this regard, consideration should be given to maximum credible event identification. This would entail identification of a set of plausible environmental factors that, subject to any given exposure profile, would result in maximum potential for irreversible adverse effects.

Variability among the cosm data

The Panel observed that there appears to be considerable uncertainty regarding the estimate of the PATI-LOC as a consequence of variability among the cosm data. This was illustrated in presentations made on behalf of Syngenta during the meeting. The Panel presented several means of characterizing this uncertainty. One general approach would be to use a bootstrap method to quantify the uncertainty inherent in the data, e.g., sampling distinct subsets of the data and estimating the PATI-LOC for each subset and then quantify the uncertainty among estimates or sample with replacement and account for correlation among subsets. Another option would be to use a series of sequential random samplings without replacement of the cosm data to examine if there is a convergence in PATI-LOC estimates with increasing cosm data. For example, first randomly pick a minimal cosm data set among the available data as required to perform logistic regression, estimate a PATI-LOC for this subset, then randomly select, without replacement, another datum from the cosm data and add this to the initial random set and lastly, estimate the PATI-LOC based on these data. Repeat this sequential random sampling multiple times and plot the resulting average and standard deviation of the estimated PATI-LOC values as a function of the number of cosm data. This will provide a measure of whether the PATI-LOC is converging to a fixed value and provide insight into the utility of additional cosm data. Another approach would be to group the data by exposure duration (e.g., estimate the PATI-LOC using the cosm data with exposure durations of 20 days or less), and then sequentially add subsequent cosm data with increasing exposure durations to evaluate the influence of cosm exposure durations on the PATI-LOC estimate. The Panel noted that the latter two methods are not as statistically rigorous as the bootstrap method. The Panel was concerned that the single-species toxicity tests (Stage 2 of the four-stage process, Fig. 15, p. 100 of the White Paper) were based largely on species typically observed to be "tolerant" to various forms of anthropogenic stress, which could result in underestimating the effects of atrazine in natural plant communities. Although these species may be common members of aquatic plant communities that tend to be found in atrazine-use areas, they may simply be better survivors in such places. Such a possibility may call into question how representative these species are of the range of potential sensitivities of "all aquatic plants" to atrazine (see Fig. 15). Ideally, a variety of species that would be more representative of unperturbed "reference" sites (see related response to charge question 6), should be used because they are more generally sensitive to chemical stressors. Therefore, the Panel recommended that a broader range of species sensitivities to atrazine be included in such tests to better

represent the potential effects to a full of range of primary producer species in the system and maintenance of ecosystem integrity.

The Panel also had concerns about the degree of intraspecific variation in some of the SGR EC_{50} values used in deriving the PATI relationship. When multiple studies were conducted using the same species, very large variability in SGR EC_{50}'s was sometimes observed, e.g., *Pseudokirchneriella subcapitata* (formally known as *Selenastrum capricornutum*) exhibited a nearly 5-fold range, from 50-236 µg/L (Appendix E, Table 1, p. 16). This high degree of variability calls into question the comparability and/or quality of at least some of the data used, and furthermore, leaves open the possibility that different strains may have different exposure histories resulting in atrazine resistance in some of them. The Panel recommended that study-acceptability criteria be incorporated in the screening process to ensure that some meaningful portion of the studies used in the development of the CE-LOC include species known to occur in reference sites where atrazine is absent, as opposed to focusing largely on potentially tolerant species, or species common in generally disturbed areas, and are typically cultured in laboratory settings. Studies should provide support that the strains used have not been exposed to atrazine (i.e., they should come from previously unexposed, wild populations) in order to establish the representativeness of the observed toxicity response for a given species. If these additional criteria are not met by the studies used, then the CE-LOC should be qualified in a way that clearly reflects that it is not known whether the CE-LOC represents the full range of primary producer sensitivity to atrazine.

Question 8. The 2009 SAP recommended using an effects index or concentration metric, rather than categorical LOC thresholds in order to take advantage of data from Syngenta's Atrazine Ecological Exposure Monitoring Program (AEEMP). At that time the LOC threshold for atrazine effects to plant communities was established at 10 µg/L for a 60-day rolling average. The current analysis using the Plant Assemblage Toxicity Index (PATI) indicates the CE-LOC can range from 4 to 7 µg/L (Chapter IV, Section 14.3 & 14.4). *Please comment on this CE-LOC and whether it reasonably represents a range below which permanent or irreversible change in aquatic plant community structure, function, and/or productivity due to atrazine exposure would not be expected.*

Panel Response

The Panel appreciated the PATI approach as a thoughtful method for deriving a CE-LOC from a wide range of cosm and individual photoautotrophic species toxicity tests. Based on the AEEMP chemographs, 60 days seems to be a robust interval for consideration of exposures. The majority of the Panel indicated that the 60-day rolling average of 4 to 7 µg/L CE-LOC proposed by the Agency represents a level below which no permanent or irreversible change in aquatic plant community structure and/or productivity will occur, based on the primary producer data shown. Having said this, the Panel could not validate the 4 to 7 µg/L CE-LOC due to a number of concerns with the LOC methodology previously described in responses to charge questions 6 and 7. As a result, the Panel expressed minimal confidence in the calculated CE-LOC.

The Panel provided the following specific comments to supplement those previously made in response to the two previous questions.

1) Each step in the multi-step LOC methodology is associated with inherent error, and then propagated along each step so that the accumulated error will likely be quite large. The Panel stated that there needed to be a clear idea of the magnitude of the propagated error before any discussion of the precision and accuracy of the estimate.

2) There is a lack of a 95% C.I. associated with the CE-LOC estimation. In other words, is it reasonable to propose the CE-LOC to be 4 to 7 µg/L while not knowing if this number is the lower 95% C.I. for conservatism or if it is the central tendency of the estimate? The Panel indicated that the "true" range of the atrazine CE-LOC was needed before a meaningful comment on the range could be made.

3) The Panel had concerns with the selection process of the final cosm dataset discussed in response to charge question 6. In brief, the Panel provided examples that many mesocosm effect scores were incorrect or unjustified at the level of ≤ 20 µg/L due to misinterpretation of the data where effects were claimed when, in actuality, they did not occur. The same point was made by Dr. Giddings (on behalf of Syngenta) during the public comment period (a copy of the presentation is found in the public docket, EPA-HQ-OPP-2012-0230). Revising the scoring of the cosm dataset will change which studies will be considered in estimating the CE-LOC. In all likelihood, the recalculated CE-LOC will be well above the EPA current level of 4-7 µg/L. However, it should be noted that this effect could be counteracted by future cosm studies using a dataset comprised of primary producers exhibiting a broader range of sensitivity to anthropogenic stress (i.e., incorporating the "reference" concept discussed under question 6), which may have the effect of lowering the CE-LOC determination.

Question 9. Based on previous analyses of the available ecotoxicity data, EPA concluded for atrazine that the level of concern for effects on aquatic plant communities (CE-LOC) was lower than the atrazine concentrations observed to produce significant direct or indirect effects on invertebrates, fish and amphibians. Given the current analysis of the ecotoxicity data (Chapter I, Section 6) and the Aquatic Plant Community LOC methodology, EPA continues to believe the original conclusion still holds true. *Please comment on how well the available database supports EPA's conclusion that the CE-LOC is lower than exposures that result in significant effects on the growth, survival and reproduction of aquatic animals.*

Panel Response

Further testing for effects of long-term, low concentration exposures of atrazine

The Panel stated that without further testing of the effects of long-term, low concentration exposures of atrazine and its degradation products, EPA's conclusion (stated in the charge question) that the *"CE-LOC is lower than exposures that result in significant effects on the growth, survival and reproduction of aquatic animal"* cannot be supported. Exposures include duration and persistence through repeated applications of atrazine. The Panel noted that a regression design could cover a range of environmentally relevant exposures without being constrained by an estimated CE-LOC.

As of the meeting, there are no published data indicating that environmentally relevant concentrations of atrazine at or below 4 to 7 µg/L are directly or acutely toxic to invertebrates, fish or amphibians (e.g., see analysis by Rohr and McCoy 2010), although there are some published studies that suggest that atrazine exposures could lead to direct behavioral changes (e.g., hyperactivity), a cascade of effects (e.g., population loss due to decreased food, or endocrine system alterations leading to impaired reproduction and development), or changes in immune response (e.g., reduced ability to withstand indigenous pathogens). In addition, panel members noted that there was compelling laboratory evidence to demonstrate that low levels of atrazine could impact reproductive endpoints in fish (Tillit et al. 2010) and in amphibians (Hayes et al. 2002, 2003, 2006; Storrs and Semlitsch, 2009). Natural field studies

(Hayes et al. 2003) also provide evidence that low levels of atrazine could affect reproductive endpoint in amphibians. The Panel supported further investigation of the significance of possible reproductive effects of atrazine within populations, between populations, and across species (e.g., Alvarez and Fulman 2005, Rohr et al. 2006, 2008; Storrs and Kiesecker 2004).

Additional studies need to examine (but not be limited to): 1) the susceptibility of amphibians to parasites and cascade of effects (e.g., Rohr et al. 2008; Baxter et al. 2011; Koprivnikar et al. 2007; Kiesecker 2002), 2) immunosuppression in amphibians (e.g., Brodkin et al. 2007, Rohr and McCoy 2010, Forson and Storfer 2006), and 3) alteration of amphibian gonadal development (SAP 2008). For example, few immunological studies have looked at environmentally relevant exposures of atrazine; however, the small amount of work that has been done on amphibians suggests that atrazine can act as an immunosuppressant. At concentrations of 3 and 30 µg/L, atrazine significantly decreased circulating eosinophils in wood frogs (*Lithobates* [*=Rana] sylvatica*) (Kiesecker 2002).

The Panel indicated there were insufficient data on indirect effects of atrazine on herbivores (wherein herbivores would be impacted not by direct toxicity of atrazine, but by detrimental impacts on their photoautotrophic food source) and the effects of resistance evolution and tolerance to atrazine on aquatic plant communities. The Panel postulated that if plants were damaged by exposure to atrazine then there would be concomitant adverse effects on herbivores and subsequently ecosystem structure and function.

Mixtures and degradates

The EPA's acceptance criteria for cosm studies (Appendix D, p. 3) indicated that tests must involve administration of atrazine only and not mixtures or multi-active ingredients. The Panel pointed out that isolation of atrazine for toxicity testing and cosm studies likely render direct assessment of atrazine effects much more straightforward in the laboratory. However, it would be flawed in terms of how atrazine affects growth, survival, and reproduction of aquatic organisms under "realistic conditions" because atrazine will be only one of the many stressors in the environment which could act additively or synergistically with atrazine to heighten the effective toxicity of a given measured concentration of atrazine. Therefore, likely interactions among environmental stressors with atrazine should be considered, or at least not eliminated from the analysis when available in the published literature, in the process of establishing meaningful effects levels. The section on "Stressor of Concern" (Chapter 1, Section 5; pp. 29-30) in EPA's White Paper does explain that while *"quantitatively predicting the combined effects of multiple variables on mixture toxicity to any given taxa with confidence is beyond the capabilities of the available data and methodologies", "available toxicity data for environmental mixtures of atrazine with other pesticides will be presented as part of the ecological risk assessment" and "a qualitative discussion of implications of the available pesticide mixture effects data on the confidence of risk assessment conclusions will be addressed as part of the uncertainty analysis."*

The "Stressor of Concern" section in the White Paper also contains the EPA's rationale for not including atrazine degradates in the analysis, i.e., atrazine is more toxic and found in higher concentrations than any of its degradates; therefore, the CE-LOC calculated for the parent compound should be considered protective of any effects caused by the degradates. The Panel pointed out that this explanation may overlook the possibility that a given measured concentration of atrazine in the presence of its (unmeasured) degradates may be more toxic than that same concentration of atrazine in the absence of degradates. The atrazine concentration alone, without consideration of degradates, is used for the calculation of the daily PATI values leading to the determination of the EEF. Therefore, any toxicity contributed by atrazine degradates would be unaccounted in EEF. To the extent that atrazine in combination with its degradates results in toxic effects exceeding those that would be anticipated from monitoring of, and accounting for, atrazine alone, omission may result in underestimating risk.

Question 10. *Please comment on the strengths and limitations of EPA's development and use of bias factors (Chapter V, Section 16.1) for addressing uncertainties in monitoring data.*

Panel Response

The Panel commented on the strengths and weaknesses with the use of bias factors to address uncertainties in the monitoring data and suggested a number of clarifications and refinements that should improve both the acceptability and understanding of the bias factor approach. Many of these were previously discussed during the 2011 SAP meeting (SAP 2011).

1) Clarify why the specific data used in the analysis were selected.

2) Describe with examples and in detail how data in-filling was performed.

3) Describe with examples and in detail how these data were used in the regression step.

4) Include sites representing all types of streams. Sites should not be limited to streams that have present and/or future use as a community water system (CWS) source as was recommend by the 2011 SAP relative to human-health significance.

5) Describe how the uncertainty in the estimation of the bias factor will be factored into the likelihood of the site exceeding the LOC, the statistics of primary interest at the screening step question.

As indicated in the White Paper, observed bias factors were computed using a Monte Carlo stratified random re-sampling of the intensive single-day sampling time series of the seven site-year combinations (MO-05 2008, MO-05b 2008, MO-02 2008, MO-02 2009, MO-04a 2008, IN-11 2008), SANDUSKY 1995 listed in Table 24, p. 136) to simulate sampling time series with observations every 4^{th}, 7^{th}, 14^{th} or 28^{th} day. Each of the 10000 realizations of these time series were interpolated to produce 10000 simulated chemographs which were then processed into 10000 maximum 60-day mean atrazine concentrations. From the distribution generated, the 5^{th} percentile of the annual maximum 60-day rolling mean atrazine concentrations was compared to the true maximum 60-day rolling mean atrazine concentration to compute the bias factor for that site and sampling frequency. This produces seven conservatively high (because the 5^{th} percentile was used) bias factor estimates for each sampling interval.

The Panel noted that there is nothing to prevent sampling every 3^{rd}, 5^{th}, 9^{th} or 21^{st} day for that matter. Since sampling interval is later translated into number of samples in the 60-day period, one could conceivably have scenarios of 2, 3, and 4, to say 30 samples in the 60-day period. If one has enough data, it is also possible that this exercise might also be attempted for sites that address a wide range of basin sizes, average flows, site types (CWS reservoir, flowing stream, etc).

With seven bias factors for each sampling interval, the next step is to come up with one bias factor for each interval. By grouping or regressing the bias factors against sampling interval, it is possible to establish one bias factor that has some probability of coverage associated with it (i.e., could be said to cover 95% of potential bias factors at this sampling interval).

Application of bias factors

As noted in the White Paper and during the meeting discussions, analysis of bias factors based on the selected seven site-year combinations (see Table 24, p. 136) is intended as a demonstration analysis to illustrate the proposed approach, particularly focusing on the regression method (discussed further in Question 13), rather than a final analysis for application. As correctly stated in the White Paper, these seven site-year combinations are rather limited in geographic and hydrologic variability. The estimation of bias factors is to be part of an initial screening level step to identify sites for further evaluation and possibly increased monitoring. Bias factors estimated for these seven site-year combinations will not have a high level of precision, but this is unnecessary to demonstrate the approach.

The development of bias factors in this screening level analysis is based on selected high sampling intensity monitoring data from the AEEMP and NCWQR monitoring programs. The purpose of the bias factor adjustment is to address the expected underestimation of the true maximum 60-day mean concentration for streams with limited sampling. As expected, the application of bias factors to adjust the maximum 60-day means will decrease the likelihood of underestimating the true value in some cases, but at the same time, will increase the likelihood of overestimation in other cases.

The Panel summarized the EPA's justification of the 60-day period reported in the Appendix E, Section 4.2.1 of the White Paper. The 60-day period is the longest cosm treatment with no effect, and almost all periods of significant exposure in the AEEMP monitoring data are less than 60-days. Beyond 60-days, there are limited cosm data. The median cosm duration is closer to 30 days and the mean is 57 days pointing to the impact of the six studies with durations >300 days (p. 48, presentation by Drs. Brain and Nair representing Syngenta, "Overview and Evaluation of EPA's Current Methodology for Developing Levels of Concern," available in public docket, EPA-HQ-OPP-2012-0230). The EPA noted in the White Paper that using an averaging period of less than 60-days leads to some invalid conclusions that less exposure is needed to elicit exposures than are actually involved.

The Panel concurred with recommendations provided by the 2011 SAP and incorporated additional comments on the application of a bias factor to exposure statistics and approaches on how to apply site-based bias factors (p. 22, SAP 2011).

> *Application of a bias factor to exposure statistics calculated from simple linear interpolation of sparse monitoring data is a potential simple and practical approach to evaluating data from a variety of monitoring frequencies to get either unbiased or conservatively high biased preliminary estimates of exposure metrics, depending on how the factor is derived. The approach is primarily applicable to sites with moderate frequency monitoring data, such as weekly or biweekly, so that initial biased sample estimates are more or less in statistical control. Quarterly data, for example, would be too sparse to use for short duration sample estimates*.

The Panel provided three potential approaches (originally discussed by the 2011 SAP) that could be used to address the problem of how to apply site-based bias factors determined for intensively monitored sites to other sites with sparse monitoring data. These are summarized as follows:

1) **Evaluation of "homogeneous" groups to develop a categorical system of bias factors.** This is the approach referred to in the charge question. If there are usable discreet groups, as opposed to a continuum, then perhaps 30 sites in each group with 10 years of high quality data would be a reasonable place to start. The Panel acknowledged that reservoirs, which account for a large

proportion of the CWS's, will probably be very difficult to categorize using this approach because of their highly variable characteristics, such as volume and residence time.

2) **Regression of bias vs. explanatory variables, such as basin characteristics and water-body type, to express bias as a continuum governed by specific basin characteristics.** This approach could be promising to address at least certain parts of the problem, such as watershed size for flowing streams, but more data for multiple years would be needed at selected sites, as well as at additional sites with intermediate basin sizes.

3) **Use characteristics of a "worst case group", such as small basins, to yield a conservatively high bias factor for protective screening that would trigger monitoring.** This is a practical approach that can be used now for CWS's because there is relative confidence that flowing water sites with small basins, such as the AEEMP sites and other small basin sites, define the worst case bias factors for both larger flowing streams and reservoirs. There are a significant number of CWS's within the watershed size range of the AEEMP sites which could use the AAEMP characteristics to define small-basin CWS sites in vulnerable settings.

Strengths and limitations of the bias factor approach

The Panel listed the following strengths and limitations of the bias factor approach.

Strengths:

1) Establishing the operational definition of bias as the ratio of the 5th percentile of the distribution of the max 60-day moving averages distribution from the Monte Carlo study to the true value for each site, provides the statistical means to make a probabilistic statement about coverage. That is, we would expect this bias factor to be exceeded in only 1 in 20 "years."

2) Limited focus on the 2nd and 3rd quarters of the year because these periods are most associated with high atrazine use in the field.

3) Limited the sites used to those with near complete daily data, which would require low in-filling (interpolation).

Limitations:

1) Uses only a few calibration site years, i.e., seven site year combinations used, which are limited in geographic and hydrologic variability. The Panel acknowledged that the Agency's analysis of bias factors based on the seven sites is intended as a demonstration analysis to illustrate the proposed approach, particularly focusing on the regression method (discussed further in charge question 13), rather than a final analysis for application.

2) Uses a simple stair-step interpolation method to in-fill missing data for sparsely sampled time series. See discussion of "in-filling" below.

3) Does not take into account the monitoring method (grab vs. autosample) or the number of samples taken in a day.

4) Only four sampling intervals were considered; whereas, there might be as many as 28 possible sampling intervals that could be used in the simulation.

5) Limits on the range of basin sizes, average flows, and site types represented by the data constrains an assessment of how these factors might contribute to the calculation of bias factors.

6) Bias factors were applied to watersheds and regions that were much different than the watersheds from which they were calculated;.

7) The Monte Carlo analysis does not take into account "serial correlation" or the dynamic nature of rainfall events which could be important in situations where sampling is less frequent than every 4th day.

Estimation of bias factors from the seven site year combinations

As indicated in the White Paper, observed bias factors were computed using Monte Carlo stratified random re-sampling of the intensive single-day sampling time series of the seven site-year combinations down to time series with observations every 4th, 7th, 14th or 28th day. Each of the 10000 realizations of these time series were interpolated to produce 10000 simulated chemographs which were then processed into 10000 maximum 60-day mean atrazine concentrations. From the distribution generated, the 5th percentile of the annual maximum 60-day rolling mean atrazine concentrations was compared to the true maximum 60-day rolling mean atrazine concentration to compute the bias factor for that site and sampling frequency. This produces seven conservatively high (because the 5th percentile was used) bias factor estimates for each sampling interval.

As noted previously by the Panel, there is nothing that keeps one from sampling every 3rd, 5th, 9th or 21st day. Since sampling interval is later translated into number of samples in the 60-day period, one could conceivably have scenarios of 2, 3, 4 or even 30 samples in the 60-day period. If there is enough data, then it might be possible for this exercise to be performed at other sites to address a wide range of basin sizes, average flows, site types (e.g., CWS reservoir, flowing stream, etc).

For each sampling frequency, the preceding protocol produces seven bias factors; one for each site-year, the next step is to come up with one bias factor for each interval. By grouping or regressing the bias factors against sampling interval, it is possible to establish one bias factor that has some probability of coverage of 95% of the potential bias factors at this sampling interval.

The Panel agreed that true bias always tends to be in the direction of underestimation of atrazine exposures because short-lived high concentration events have a higher likelihood of being missed with sparser sampling. In non-worst case site/years, there is less of a tendency for extreme events to occur, which will result in more subdued hydrographs and subdued chemographs, although this is not guaranteed. Hence, even with sparse sampling the bias adjustment is expected to be smaller than during years with more extreme rainfall/runoff events.

The Agency currently estimates the bias factor directly as the ratio of the true maximum 60-day rolling average to the corresponding estimate from a Monte Carlo simulated chemograph. An alternative, which reduces uncertainty in bias factor estimation by explicitly accounting for the stratified random sampling paradigm, would be to estimate the bias factor as a function of sample interval. Consider that for a specified sampling frequency (e.g., every 4th day), the Monte Carlo re-sampling approach used by the Agency selects "sampling" days randomly from within each sample interval. The re-sampling is repeatedly applied to the 180 day time series (2nd and 3rd quarters) for each site year, identifying a unique set of samples and sample intervals (e.g., for a 4 day sampling frequency, 45 days and 44 intervals, respectively) for each of the 10000 simulated chemographs. For a given *n* day sampling

frequency this protocol results in a distribution of sampling intervals, as outlined in the following table (**Table 3**) for a 4 day sampling frequency.

Table 3. Stratified Random Sampling with 4 Day Frequency: Sampling Intervals and Frequencies							
Days between Consecutive Samples	1	2	3	4	5	6	7
Relative Frequency of Occurrence	6.25%	12.5%	18.75%	25%	18.75%	12.5%	6.25%
Average number of sample intervals	27500	55000	82500	110000	82500	55000	27500

In-filling within each sample interval is used to create a daily time series for each simulated chemograph (e.g., linear interpolation to date, additional in-filling options are described below). The relative accuracy of this in-filling process will vary with sample interval, decreasing as sample interval increases. For a given sampling frequency, a temporal resolution "smearing" that results from the stratified random sampling which is manifest as generally greater bias factors compared to systematic sampling. A possible means of reducing this impact is to calculate an interval specific "bias" factor. This would involve calculating the ratio of the "true" average concentration to the corresponding simulated average value for each sample interval. For the 10000 simulated chemographs with a 4 day sampling frequency there would be 55000 "2 day sample interval" bias factor estimates, 82500 "3 day sample interval" bias factor estimates, etc. (Note: The 1 day sample interval bias factor is equal to 1 by definition.) In general, it is expected that the bias factors so estimated would increase within increasing sample interval. Thus the 2 day sampling interval average bias factor would be smaller than the 3 day sampling interval average; the latter would be smaller than the 4 day sampling interval average, etc. The 5th percentile for each sample interval could thus be calculated, resulting in seven "5th percentile" bias factors for a 4 day sampling frequency (i.e., corresponding to 1, 2, 3, 4, 5, 6, and 7 day sampling intervals).

The sampling interval bias factors could be estimated using systematic sampling. The latter would provide 180 *"m* sampling interval bias factors" for an *m* day sampling interval (e.g., 176 bias factors for a 4 day sampling interval). Again the 5th percentile bias factor could be chosen. These 5th percentile bias factors could be applied to the appropriate sampling intervals for each Monte Carlo simulated chemograph and the resulting maximum 60-day rolling average calculated.

It is insightful to examine the relationship between the sampling interval (SI) bias factor (BF) approach and that followed by the Agency. For instance, when the SIBFs are estimated using only the 60 samples associated with the maximum 60-day rolling average of the intensely sampled time series, the 4 day stratified random sampling (SRS) BF calculated by the Agency will converge in principle (assuming no systematic errors with in-filling) to the following linear combination of SIBFs:

$$SRS\ BF_{4\ day}^{60\ day} = \left[0.0625 \times \left(SIBF_{1\ day} + SIBF_{7\ day}\right) + 0.125 \times \left(SIBF_{2\ day} + SIBF_{6\ day}\right) + 0.1875 \times \left(SIBF_{3\ day} + SIBF_{5\ day}\right) + 0.25 \times SIBF_{4day}\right]$$

Similarly, each SIBF, when estimated with these same 60 samples, is equal to the corresponding BF calculated using the Systematic Sampling (SS) protocol.

There are several advantages of working directly with the SIBFs which are outlined here. For instance, the SIBF approach:

1) Avoids the temporal smearing associated with the Monte Carlo SRS approach.

2) Provides a convenient means of accounting for the distinction between daily grab and composite samples (e.g., a bias factor should be applied to the daily grab samples as outlined in response to charge question 11b.

3) Allows the user to explicitly identify the temporal range of interest. The present method of deriving the bias factors with regard to the maximum 60-day rolling average concentration restricts the analysis to the portion of the time series with maximum 60-day rolling average (i.e., the interval most relevant to the CE-LOC). However, it is not clear that this advantage outweighs the disadvantage of ignoring the information regarding sampling frequency and bias that is contained within the remainder of the sampled time series.

4) Provides a more accurate and conceptually consistent bias factor estimate for sparsely sampled chemographs in which the sampling frequency is variable.

5) Can be used to provide a direct estimate of the SRF BFs without the requirement for Monte Carlo re-sampling.

6) Can be used to examine the presence of a temporal trend in the sampling interval bias factors that reflects a general seasonal decrease in atrazine concentration variability.

A disadvantage of the SIBF approach is that the percentile cutoff required ensuring the identification of the 5^{th} percentile 60-day maximum rolling average is not known a priori and may require a trial and error evaluation.

In-filling for missing data

Another approach to in-filling sparsely sampled chemographs (concentration time series) and for reducing the bias factor uncertainty is to use the stream level record as a predictor in the regression equation. For the AEEMP site-years, the stream level for each site-year was continuously recorded with pressure transducers and data logging. This record provides a surrogate for flow rate (and may be calibrated to the latter) and identifies the onset, intensity, and duration of flow events, as well as periods of base flow. If this stream level record is not available, it can be estimated for basins without gauge data using watershed characteristics and climate variables (see the response to charge question 12). A temporal (possibly linear) trend, and associated prediction errors in base flow atrazine concentrations can be used for such in-filling and also used to reduce uncertainty when in-filling the simulated sparse systematic sampling time-series used for bias factor estimations. This approach would be an improvement over linear interpolation, which does not account for whether or not the observation occurred during a flow event.

Flow event in-filling for both the dense annual time series and the simulated sparse systematic sampling time-series is much more challenging and of much greater importance, as these are periods during which

the greatest atrazine loads and concentrations are typically observed. For a given flow event, there is a strong relationship between the flow rate and atrazine concentration (e.g., Pappas and Huang 2008, Zanardo et al. 2012). Atrazine concentrations are also a function of timing between pre-plant atrazine applications, soil/landscape/management factors, and rainfall. In-filling would take advantage of these flow event characteristics.

Linear interpolation is a reasonable approach to in-filling missing hydrograph and atrazine concentration data where frequent but incomplete sampling ("dense sampling") is available for flow events. This should provide a first-order approximation to the true atrazine concentration time series. With samples widely spaced in time ("sparse sampling"), it is likely that some if not all of the very "flashy" runoff events will be missed, with the result that the within-event variations in atrazine concentration (e.g., peak concentration), are poorly, if at all, characterized. For chemographs produced with sparse, systematic sampling, estimating atrazine concentrations for flow events during which no samples have been collected poses an especially challenging task.

The PRZM-hybrid approach presented by Dr. Paul Hendley, Syngenta, (presentation available in the public docket, EPA-HQ-OPP-2012-0230) provides a daily estimate for such situations, but this approach does not adequately capture the flow dynamics (e.g., daily time step, crude watershed conveyance estimates). If these "edge-of-field" estimates were coupled with better watershed hydrology measures/predictions (discussed further under charge question 12), then this information might provide a basis for improving in-filling during flow events. Zanardo et al. (2012) demonstrated an approach for predicting (i.e., without calibration) atrazine flux in a basin with no gauge data that was shown to provide reasonable predictions of atrazine concentrations during observed flow events at multiple sites across several years.

Possible limitations of using bias factors with small watersheds

The Panel stated that the bias factor approach was more applicable to larger spatial scales. They indicated there were limitations to using the bias factors to interpret monitoring data within very small watersheds, i.e., those smaller than most AEEMP watersheds. Spatial variability in streams increases considerably as watershed area decreases. As one moves downstream, water quality reflects dilution and mixing processes that average flows across a myriad of landscape features, management practices, climate variables, etc. Watershed scale thus impacts the observed temporal and spatial distribution of atrazine concentrations. The literature suggests that, within a landscape association, there may be a threshold spatial scale above which averaging mechanisms are sufficient to minimize concerns regarding within-stream longitudinal variability in atrazine concentrations (Temnerud and Bishop 2005). These authors noted that spatial variations in water quality indices vary greatly within small, < 15 km^2, headwater watersheds; whereas, such indices are relatively uniform above this scale. With these limitations in mind, the Panel thought it would be very challenging to use bias factors to identify small spatial scale areas of concern using monitoring data. Whether this issue is significant or not could be determined from the estimated fraction of the total U.S. atrazine use area and how many stream miles would be considered small headwater watersheds.

Question 11. Prediction of bias factors is dependent on the selection of an appropriate model. EPA illustrated (Chapter V, Section 16.1) both categorical and regression methods for prediction of bias factors based solely on the number of samples taken in the 2nd and 3rd quarters of the year (April 1st to September 30th).

a. Please comment on EPA's prediction of bias factors from monitoring data using categorical or regression method approaches.

Panel Response

The Agency acknowledged that the bias factor prediction methods are preliminary estimates based on only seven site year combinations. The Panel emphasized the importance of expanding the data to provide a more representative assessment for the regions of interest. Albeit preliminary estimates, the Panel enumerated several issues with the regression (three) and categorical (one) approaches used to select bias factors. The Panel recommended that bias factors be estimated for other sampling intervals, which could be integrated into the bias factor prediction model.

Categorical approach

The Panel stated that the categorical approach is the simplest approach. It uses the upper 95% confidence interval of the mean bias factor for the different sampling frequency classes: 25 to 46 samples bias factor of 1.87 (rounded to 2); 12 to 26 samples bias factor of 2.94 (rounded to 3); 7 to 13 samples bias factor of 4.01(rounded to 4); and ≤ 6 samples bias factor of 10.83 (rounded to 11). The Panel indicated there were two key questions regarding this approach to be answered:

1) For sites with greater than 46 samples (from a possible 180 days), will the assigned bias factor be set at 1; and assumed that the maximum or near maximum flow event was observed?

2) Why shouldn't other sampling intervals (number of samples) be used in the Monte Carlo simulation to produce "fuller bins" or more detailed bins?

Stepwise regression

Three regression approaches were used for illustration purposes only. All three approaches used only one explanatory variable and also combined the number of samples in the 2nd and 3rd quarter. Neither the watershed size nor the average flow values were significant from the stepwise regression analysis performed.

The Panel enumerated several issues with the stepwise regression approach:

1) The question as formulated to the Panel indicates that forward stepwise regression was used. Because only three factors (number of samples, watershed size, and average flow) were explored as predictors, it was feasible to examine instead models with all possible combinations of these three factors. Forward stepwise regression may have missed the "best" model. If only a forward stepwise regression approach was used as specified in the question formulation, it is possible that the "best" regression model was actually not observed. With three factors, models incorporating all possible combinations of variables could have easily been examined. For example, number of samples might be the best single predictor, but the combination of watershed size and average flow might together have produced a better model, only with forward selection that model would not be examined.

2) In the stepwise regression model building protocol, the factor selection approach used a 5% significance test to determine whether a potential predictor should remain in the model. With so little data and few factors, a higher level of significance could be used to avoid dropping factors too early in the process. Of course, with more data the 5% significance level for dropping a predictor from the model is a reasonable choice.

3) Watershed size and average flow may be important predictors of bias factors, but were not significant in this exercise because the ranges of these variables were severely limited. The screening level exercise was limited to five monitoring sites (7 site-year combinations, some sites have two years of data) which limited the range of watershed sizes and average flow represented. Such small datasets produce more uncertainty in the estimation of the regression coefficient, which leads to a greater probability of rejecting any given factor as a useful predictor. It was almost preordained that the number of samples would be the best since this is the one factor that can be controlled in the Monte Carlo study. The significance of an explanatory (X) factor in a multiple regression is somewhat determined by the range of values that factor can have (i.e. by the Var(X)). Smaller range produces more uncertainty in the estimation of the regression coefficient, leading to greater probability of rejecting the factor as a useful predictor. Watershed size and average flow may be important predictors, but were not significant here because the ranges of these variables were severely limited.

4) The White Paper describes the estimation of the bias factor using a line fit to the upper 95% confidence interval estimates (see Table 25, p.141). However, the computer programs and output presented in Appendix K (Bias Factor Regression Equations) seem to indicate that the regression was performed on the 95% prediction interval estimates. Since the estimate is meant to be applied to a single site, statistical theory suggests that the 95% prediction interval estimates should be used. This distinction should be made in the White Paper.

Regression one, two, and three

The Panel offered the following comments on each of the regression equations.

Regression one: The best least-squares non-linear regression fit to the mean of four groups of data (Figure 31,White Paper) was a four-parameter exponential decay model represented by the following equation: [$y = 470.8*exp(-0.071*x) + -458.8*exp(-0.071*x)$; $r^2=0.33$; $p<0.02$], where y is the bias factor and x is the number of samples. Note that the coefficient for the two exponential components are the same and the intercept coefficients are about the same, but different in sign. The Panel commented that the parameters of this fitted model will have poor statistical properties (commonly referred to as a poorly parameterized model). The form of the equation describing the upper 95th percentile confidence interval (of the mean) was also provided:
[$y = 1383.1*exp(-1.07*x) + 11.6*exp(-0.04*x)$; $r^2=1.0$], which is also close to a simple exponential curve since the first exponent parameter is much smaller than the second exponent parameter.

Regression two: For regression two, four groups were created, indexed by the "number of samples" used in the re-sampling study. The White Paper gave no name to these groups, but plots were labeled with "number of samples". Regression two is a regression fit to the average bias factor in each of the four sample groups. A four-parameter exponential decay model represented by the following equation: [$y = 170.5*exp(-0.56*x) + 4.1*exp(-0.02*x)$; $r^2=1.0$] was used, but because this model has four parameters to be estimated and there are only four data points being fit, a perfect fit is observed. Again, this is essentially a smooth interpolation line that is likely to look quite exponential.

Regression three: Regression three uses a 1st order linear model on log-transformed bias factor values stated as: [$log(y) = 0.79 - 0.014*X$, ($r^2=0.46$, $p<0.0001$]. The degree of departure from linearity in the data presented in Figure 32 or from the output in Appendix K cannot be easily determined, but if there is non-linearity, then the coefficient is small. The EPA states in the White Paper that the regression model for the 95% confidence interval is $log(y) = 0.95 - 0.008*X$, which is linear. The Panel indicated that this

must be incorrect because Figure 33 (White Paper) shows an appropriately hyperbolic shape for the 95% confidence bounds and even a rough straight line along this upper bound (in red) would have to have a zero intercept closer to 9.4 not 0.94.

Panel members, who were also part of the 2011 SAP, made the same key recommendation that there needs to be an adequate number of sites and years of data for this analysis. The final results of the model building process (bias factor regression equations) are very sensitive to the characteristics of the data actually used in the analysis. Ideally, the method should be applied to a model development data set in the range of 50-100 site years or more.

b. Please comment on any additional methods for estimating bias factors that would be useful in this situation.

Panel Response

The Panel provided examples of several strategies for selecting which sites to use for further development of the bias factor approach. These included models such as WARP, Przm and SEAWAVE, to produce more "complete time series" that could be used in the development of bias factors. These strategies were originally discussed by the 2011 SAP (see report, SAP 2011).
The Panel noted that the 2011 SAP envisioned providing a bias factor adjustment for each site/year based on multiple years of data at each site (SAP 2011). Realizing that these data are not available, the Panel indicated that the next option is to have the bias factor adjusted for some combination of site characteristics. The approach provided by the EPA, where the bias factor is only adjusted for the number of samples taken at the site, is the least representative of multiple sites because it uses more of a "one size fits all" approach, which is almost guaranteed to result in large bias factor estimates that are highly variable, uncertain have poor correspondence with chemograph properties, and lack of regionalization (not representative of any site/year). Using more site/year combinations will help alleviate limitations of this approach..

While the 2011 SAP recommended that the EPA "establish acceptable trade-offs between theory and practice in its assessment" (p. 33 in report, SAP 2011), the trade-off the EPA has chosen in this test case is to use only available data from a few limited site/year combinations to estimate bias factors vs. many site/year combinations. This choice results in bias factor estimates that could be quite conservative and have low associated confidence (i.e., resulting in unrepresentative atrazine concentration estimates).

Improved strategies for selecting sites

The Panel stated that to improve the confidence in the bias factor estimate (beyond that of the test examples), there should be more sites, especially sites that characterize a wider range of basin sizes, average flows, and site types. This may require using sites with less than daily sampling requiring more in-filling. The Panel discussed several strategies for selecting which sites to use for the further development of the bias factor approach.

One panelist suggested that using properly calibrated hydrological models to guide in-fill could lead to more sites for this analysis. In particular, the PRZM-Hybrid approach presented by Dr. Paul Henley, Syngenta, (presentation found in the public docket, EPA-HQ-OPP-2012-0230) seemed promising in estimating daily variability in pesticide concentrations in runoff. This approach considered two of the most important variables driving temporal variation in pesticide concentrations in runoff: rainfall (timing and spatial variability), and pesticide application (timing and spatial variability). A potential difficulty might be obtaining detailed data on the timing of atrazine application (daily, weekly, etc.). If

a complementary modeling approach was pursued, watershed best management practices could be ignored in modeling the expected variability in pesticide concentrations in runoff given the current complexity of considering mitigation practices and other best management practices. One of the public commenters (on behalf of Syngenta) outlined the potential use of a modeling approach to supplement monitoring data. This modeling approach considered the impact of management practices on atrazine runoff. However, the impact of such practices on atrazine fate could also be accounted for in other tiers of modeling as refinements are made to obtain more site specific estimates of pesticide concentrations in streams.

Another panel member suggested that it may be possible to modify the bias factor approach to account for the different classes of watersheds defined by categorizing characteristics. Conceptually, for a given number of samples, there is a distribution of bias factor estimates. This panelist suggested exploration of the extent in which variance in this distribution can be explained by site watershed characteristics.

Several panel members revisited and expanded on sampling issues discussed in the April 2009 SAP meeting report (SAP 2009). Numerous studies have shown that pesticide concentrations are strongly related to flow rate for specific flow events, as well as across flow events, e.g., for pesticides in general, see Rabiet et al. (2010) and for atrazine, see Zanardo et al. (2012). The AEEMP chemographs and flow level records suggest similar patterns. Rabiet et al. (2010) observed a strong linear relation between flow rate and pesticide concentration ($r^2 = 0.78$) over a seven month period, and an even stronger linear relation ($r^2 = 0.91$) between event and total pesticide flux (as estimated via composite sampling and event-total flow volume for a series of rainfall/runoff events) for shorter, more intense sampling periods. Given that flow level was continuously monitored for the AEEMP site-years, and the strong relationship between flow rate and pesticide concentration, there is a distinct possibility for using this information for in-filling at other sites (see discussion on "in-filling" in response to charge question 12).

Grab sampling vs. daily composite sampling

The EPA currently calculates bias factors using the assumption that the grab sample estimated 60-day average is the "true" 60-day rolling average. The Panel commented that this assumption does not reflect the within-day variation in atrazine concentration. The instantaneous flow rate is dependent on watershed physical attributes (e.g., size, permeability, antecedent moisture, conveyance factors) and the spatial distribution, frequency, intensity and duration of rainfall events which vary over the course of the year. The instantaneous atrazine concentration in a watershed is influenced by various factors including the frequency and timing of atrazine applications relative to rainfall events. Given the strong linear relationship between flow rate and atrazine concentration, it follows that a daily grab sample atrazine measurement may provide a poor estimate of the true time-weighted daily average atrazine concentration during highly variable times of the year and for certain sites that tend to vary substantially in flow rate within a given day. As an example, Rabiet et al. (2010) observed that pesticide flux varied over two orders of magnitude during a six hour flow event. For such an event, a single daily grab sample would not only have greatly misrepresented the event average concentration, it is likely that it would have missed the event entirely. This result points out that a single grab sample can differ significantly from the daily average concentration.

Rather than use the daily grab sample approach, the Panel recommended the daily composite sample approach, which will provide a more robust estimate of daily concentration. The 2008-2011 AEEMP monitoring involved auto-samplers configured to provide a composite sample (8 "sips" at 3 hour spacing per day into a common bottle) for each day. In addition, for some of these site-years, grab samples were collected every 4th day. Together, these two types of data can be used to provide an estimate of the daily

concentration variability and, carefully interpreted, may provide a basis for adjusting the grab sample estimates of 60-day rolling averages to account for within day variability. The composite measurement from the 8 sip samples corresponding to each day of the grab samples are averaged with the measurement from the corresponding grab sample to provide an average concentration for that day (day *i*) [Equation 1]. Note that the composite measurement gets more weight because it is effectively the average of the 8 sip samples.

$$Average\ Conc\ for\ the\ ith\ day = \overline{X}_i = \frac{8}{9} Composite\ Sample + \frac{1}{9} Grab\ Sample \text{ [Equation 1]}$$

Assuming that the daily variation among concentrations is independent of the size of the mean concentration, the deviations of the grab sample from the average concentration for each of the *n* days for which a 4-day sample was drawn can be used to estimate the daily variation [Equation 2].

$$Sample\ Variance\ in\ Concentration\ Within\ a\ Given\ Day = \frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X}_i - X_i^{grab}\right)^2 \text{ [Equation 2]}$$

In the preceding equation, the Panel commented that the number of grab samples for the site-year (*n)* may need to be restricted to a shorter window following atrazine application so that the data are reasonably stationary in time. The denominator for the variance estimate could be changed to reflect autocorrelation at a 4 day sampling frequency (to some value less than *n-1*, using the integral range concept). A robust estimate of the concentration variance at a time scale of 1 day could be used to relate bias factors obtained for grab sample time-series and daily composite sample time series. Such a "within day" sample variance estimate would be less robust for "flashy" watersheds in which the hydrograph and concentration time series have multiple sharp peaks and rapid recessions, but would still provide a means of estimating a confidence interval regarding the true 60-day rolling average.

One of the public commenters during the meeting indicated that states often have even more stream monitoring data than is currently available to the EPA. The Panel noted that on a scientific basis, the bias factor methodology could be applied at different geographic levels, e.g., national or state.

Question 12. EPA illustrated (Chapter V, Section 16.1) both categorical and regression methods for estimation of bias factors as a function of the sampling frequency of monitoring data. Step-wise regression analysis indicates that watershed size and average flow rate in the 2nd and 3rd quarters of the year are not significant variables for prediction of bias factors. However, the number of samples in the 2nd and 3rd quarters of the year was found to be a significant variable, accounting for 46% of the variation in the bias factor. *What other variables, if any, should be considered in the prediction of bias factors?*

Panel Response

The EPA performed a preliminary analysis based on seven site year combinations for bias factor estimation. The Panel indicated that the next step is to increase the data set to produce a regionalized (e.g., Corn Belt) estimation of bias factors (see also response to charge question 10). A comprehensive set of bias factors based on as much of the available monitoring data as possible should be developed prior to attempting to predict bias factors as a function of physically based characteristics (e.g., sampling scheme, watershed area, topography, land use classifications, hydrologic response indices, climatic factors, etc.). Several panel members, who were also on the 2011 SAP, commented that a number of the

general points made in response to this charge question were also discussed at the 2011 SAP meeting (see SAP 2011).

The Panel provided the following suggestions for improving bias factor prediction:

1) **Composite sampling.** The utility of composite sampling relative to grab sampling was cited by panel members. The simple linear interpolation approach used for in-filling may be most appropriate to in-fill between composite samples (as opposed to grab samples) for estimating bias factors. The 2008-2011 AEEMP time series involved daily composite samples (eight 3-hour interval samples composited into one daily sample). These data can be further "composited" to evaluate the influence of not only sampling frequency, but also sample compositing scheme on bias factor estimation. For example, Rabiet et al. (2010) observed that weekly composite samples were more accurate than grab or composite samples over a lesser period for estimating pesticide loadings and thus perhaps for estimating n-day rolling average concentrations. At present, each "sample" used to estimate a bias factor represents either a grab sample or a daily composite sample. However, there is no need to restrict the compositing interval to one day. The Panel suggested that the EPA examine different sampling frequencies, e.g., 4, 7, 14, or 28 days, the relative efficacy of grab, e.g., 1, 2, 4, 7, or 14 day composite samples (or, in general, a composite sample for an arbitrary number of days) with respect to estimating *n*-day rolling average concentrations and corresponding bias factor estimates. If a composite sample, derived from samples taken at 4, 7 or 14 days (no need to restrict to these compositing intervals), results in smaller bias factors and more reliable 60-day rolling average concentration estimates, then such a reduction in the magnitude and variability of the bias factors may improve the quantitative relationships between watershed physical factors and bias factors. The influence of sampling frequency in conjunction with composite sampling can be analyzed using systematic sampling (e.g., every 4, 7, 14, or 28 days, similar to what public speakers presented during the meeting) or the sampling interval bias factor approach outlined in Section 10.3 of this report.

2) **Methods for in-filling.** In addition to the sampling scheme, the variability and magnitude of bias factors is also influenced by the method chosen for in-filling between samples. Inaccuracy in the in-filling method increases bias factor uncertainty, thus weakening the opportunity to establish relationships with physical characteristics. This highlights the importance of the in-filling process as discussed in response to charge question 10. A simple linear interpolation between samples ignores much that is known about atrazine fate and transport, including the influence of application method and timing, precipitation timing, duration and intensity, evapo-transpiration, antecedent soil moisture, runoff ratios and runoff volumes, land use characteristics, land cover, topography and watershed architecture, artificial drainage, soil properties such as permeability and depth to impermeable layers, etc. The question arises as to how to best use such ancillary information to augment the in-filling within the context of creating a watershed screening tool (e.g., without developing in-filling approaches that are too complex).

   In response to charge question 10, the Panel suggested the use of water level as a regressor and noted the challenge with regard to in-filling during flow events. The PRZM-hybrid approach outlined by Miller et al. (2012) and presented by Dr. Paul Henley, Syngenta, during the meeting (presentation found in the public docket, EPA-HQ-OPP-2012-0233), may be useful for such in-filling. This modeling approach can be used to predict a watershed average daily runoff volume and corresponding atrazine concentration. As such, it can be used to predict the average stream concentration for each runoff event. Such a prediction accounts for atrazine application, sorption and dissipation, rainfall timing, duration and intensity, spatial variations in soil properties and evapo-

transpiration, etc., and may prove useful for in-filling sparse chemographs. The utility of this approach could be evaluated using the event-based model estimates as an explanatory variable for predicting event-average measured concentrations (e.g., with event-average concentration estimated using AEEMP daily composite sample chemograph time series). If this results in a regression equation that can be used to explain a significant portion of the event-based atrazine load variability between flow events within and across watersheds, then this regressor could be used for in-filling sparse chemographs by using the prediction interval as a constraint on the corresponding bias factor.

One panel member felt that the Richards' interpolation function, i.e., event concentration time series modeled as a triangular function with fixed duration (presentation by Dr. Paul Henley, Syngenta, available in the public docket, EPA-HQ-OPP-2012-0233) for AEEMP monitored site-years is inappropriate in that it does not take into account the coincidence of sample time with the hydrograph peak nor with the event flow duration. The hydrograph is continuously monitored via stream flow level for the AEEMP site-years, thus this information could guide within-event interpolation (flow level recording is a relatively inexpensive investment). Presuming one or more grab samples are available during a flow event, a triangular interpolation function for an event chemograph may prove useful (the interpolation would necessarily differ for composite samples). The tip, peak and tail time of the chemograph could be approximated by assuming coincidence with the recorded flow level estimate of the event hydrograph tip, peak and tail times. Event flow isolation may be required for overlapping rainfall/runoff events.

As time and resources permit, other options for conditioning the in-filling could be explored which capitalize on the information-rich AEEMP data. One such option would be to use a regionalized regression equation for predicting event-based atrazine concentration statistics. This could involve using the 2011 CB WARP model as a prototype and include additional explanatory variables that estimate the interaction of atrazine dissipation, application timing and temporal variability in surface runoff. The Cumulative Vulnerability Index (CVI) proposed by Lerch et al. (2011) is an example of one such variable, as is the preceding PRZM-hybrid event-based concentration estimate. A similar index could be derived from the modeling approach of Zanardo et al. (2012) which provided reasonable predictions of daily atrazine concentrations over several years in an "ungauged" watershed (i.e., the model was calibrated at the field scale in an area disjoint to the watershed). Such a regression should also include non-redundant, event-based hydrologic indices. For such indices, see the work of Olden and Poff (2003) which examined 171 hydrologic indices (HIs) and identified sets of non-redundant daily, weekly, monthly and annual indices. Yadav et al. (2007) developed regression equations to predict HI from soil, land use and climate factors and other physical characteristics of watersheds (a comprehensive list of explanatory factors is given in Tables 1 and 2 of their work while the list of HIs are given in Table 3). They employed regression predicted HI values and prediction intervals to calibrate hydrologic models and quantify uncertainty in model predictions for ungauged watersheds. Zhang et al. (2008) built on the work of Yadav et al. (2007) to provide improved model calibrations and increased accuracy in estimating hydrograph time series for ungauged basins. The HIs chosen for predicting event-based atrazine concentration statistics should also include flow rate (e.g., Rabiet et al. 2010) and the flashiness index developed by Baker et al. (2004). Such regression approaches may prove useful for directly estimating n-day average atrazine loads and concentrations.

3) **Watershed classification.** The "one-size" fits all bias factor approach presented in the White Paper results in considerable uncertainty. This uncertainty may be decreased by watershed classification. For example, the accuracy of the bias factors likely decreases as the flashiness of the watershed increases. Accuracy may also be impinged by as yet other unknown factors. Thus, it would be

useful to categorize watersheds with respect to the watershed characteristics, management practices and other environmental factors outlined by the 2007 and 2009 SAPs (SAP 2008; SAP 2009) as well as the factors and response indices outlined in the in-filling discussion in item 3 above (e.g., CVI, PRZM-hyrbid predictions, Table 3 in Yadav et al., (2007)). For example, indices related to flashiness, lag between time peak rainfall and hydrograph peak, etc. (Yadav et al. 2007). Cluster analysis or a similar statistical measure could then be used to associate distributions of bias factors with watershed classes. Classes would be classified according to similar characteristics and response indices. A watershed class specific regression would then be applicable. With such an approach, bias factors might vary between years for a given watershed due to variations in cumulative rainfall, rainfall timing, duration and intensity of rainfall events, and response indices, such as flashiness. All of these factors may lead to greater uncertainty from one year to the next with regard to estimating rolling average concentrations for sparse chemographs.

The 2009 SAP report (p. 57) states that, "*several sites have fortuitously been identified, and thus there is now an increased focus on identifying regions with shallow impervious layers in co-occurrence with sloping topography. However, the question remains, are there other watershed areas that have physical/land use characteristics that also lead to elevated exposures?"*

Zanardo et al. (2012) identifies such a watershed area, demonstrating that tile-drained watersheds can lead to significant atrazine exposures. In their work within a relatively level watershed without shallow impervious layers, observed atrazine concentrations produce 60-day rolling average values that are similar in magnitude to the proposed CE-LOC.

Question 13. EPA examined (Chapter V, Section 16.1) the performance of various regression equations to assess the failure percentage for identification of monitoring site-years with true maximum 60-day average concentrations exceeding the CE-LOC for atrazine. This analysis showed that application of a bias factor, based on sample number during the 2nd and 3rd quarter of the year, substantially reduced the number of sites with underestimation of true maximum 60-day means. *Given the EPA analysis, what other tests, if any, should be conducted to assess the performance of regression models for prediction of bias factors?*

Panel Response

The use of bias factors to substantially reduce the number of sites which underestimated the true maximum 60-day mean was expected and not a surprise to the Panel. The Panel recommended that the performance of the final models be addressed by using standard regression techniques and diagnostics to achieve the best fitting model. One can formally test the fit of a model if there are several replicate observations with the same values of the covariates. This would be a traditional goodness-of-fit test that is typically used in regression analysis.

Regression models should be evaluated with an independent data set. This testing data set should cover the full geographic and hydrographic range of watersheds to which the model is to be applied. If about 2.5% of the bias factors are above the upper 95% prediction limit, then the model is performing well. The "leave-one-out" cross validation method assures the fit of the model if the data set is not large enough to allow a separate testing data set.

The Agency has proposed combining several conservative processes to establish limits on the true maximum 60-day mean concentration. As a consequence, it is possible that the final limit may be

unrealistically conservative. Some of these conservative steps are outlined here. First, the PATI-LOC was derived based on cosm studies which did not adequately evaluate the recovery from atrazine exposures. Second, the cosms effects/no effects were very conservatively rated by the EPA (mis-scored according to the Panel as discussed in charge question 6)) and the PATI LOC is extremely sensitive to the cosm ratings. Third, the bias factor estimates employ a 5th percentile cutoff which results in a very conservative estimate of 60 day rolling average concentrations and considers the frequency of occurrence of false positives.

Question 14. An important consideration for bias factor prediction is the ability to identify monitoring sites that potentially exceed the CE-LOC for atrazine. EPA provided an example (Chapter V, Section 16.2 and 16.3) of the use of a log-linear regression equation to estimate bias factors from the USGS National Water Quality Assessment Program NAWQA data and identification of monitoring site-years exceeding the CE-LOC for atrazine. This analysis identified sites in both the Midwestern United States as well as outside the major corn growing areas with atrazine concentrations potentially above the CE-LOC. *What are the strengths and limitations of using a single regression model for prediction of spatially and temporally explicit bias factors for a nationally-distributed monitoring data set*?

Panel Response

The regression model approach illustrated in Chapter 5 of the White Paper has both strengths and limitations, and these are affected by the premises for applying the method. In answering this charge question, the following assumptions were made:

1) A CE-LOC has been determined that establishes the concentration statistic of interest. Suppose that it is the maximum 60-day rolling average during any particular year. The duration of this rolling average, whatever is selected, will have a major effect on determining what datasets qualify for model development and for model application because the bias of estimates from monitoring data depend on that duration.

2) The model has been developed using data for sites that have sufficient monitoring data to reasonably estimate "true" values of this concentration statistic for a particular site-year and that represent the range of geographic conditions and years necessary for applying the model. The importance of including a fully representative range of geography and years has been discussed in previous SAPs and remains critically important.

Strengths of single regression method:

1) Provides a consistent, uniform approach to using a wide range of existing and potentially future data sets to estimate site and year specific estimates of the maximum 60-day average for comparison to the CE-LOC.

2) Is simple and transparent.

3) Can be used to estimate the confidence bounds on the estimated bias-corrected value, which can then be used to estimate probability in exceedance of the CE-LOC and to compare to the reliability of alternative methods. However, a method that will incorporate the uncertainty in the original uncorrected estimate for the monitoring site into the model has yet to be developed.

Limitations of single regression method:

1) With more monitoring sites (and site years) and watershed characteristics, the Panel stated that the log-linear model will likely be inadequate to describe the majority of the variability in the data. More complex models, such as the multivariate regression tree model (De'ath 2002, Prasad et al. 2006), might be explored. This approach can facilitate both estimation of bias factors and identification of conditions under which the uncertainty of the resulting bias factor estimate is too high to reliably use this approach.

2) Success depends on the representativeness and completeness of the model development data set and the quality of the final model to represent those conditions. Such a development data set needs to include a much wider range of geography, size, and other characteristics to meet this requirement, and there need to be more site years for model development.

3) Applications of the model need to be constrained to data sets that meet specific data requirements for such factors as sampling intensity and seasonal coverage, such that it is sufficient to yield an adequate initial estimate to be "bias corrected." Specific decisions on such criteria need to be made.

4) The biggest overall limitation of the approach is that most stream miles for which estimates are needed have no monitoring data that will meet the minimum requirements for making an initial sample estimate to be bias-corrected. Stream miles are the appropriate geographic unit of assessment.

5) Most stream miles within atrazine-use areas will need to be addressed using an alternative approach, such as a WARP or similar empirical multivariate model. This approach has been recommended and discussed in previous SAPs.

6) For some sites with monitoring data, the geographic statistical model will have less uncertainty than the monitoring data estimate.

7) This approach would require a nationally consistent approach to identify qualifying data for monitoring site years.

The Panel also suggested that if one were to examine the probability of exceeding the CE-LOC, then the entire range of methods from sample estimates based on intensive data, to bias-corrected sparser data, to regression-based multivariate extrapolation models, could all be used and judged by the same standard. All of these methods could be part of a set of methods that provide a step-wise process for identifying and dealing with streams of potential concern.

REFERENCES

Topic A References

Baxter L.R., D.L. Moore, P.K. Sibley, K.R. Solomon, M.L. Hanson. 2011. Atrazine does not affect algal biomass or snail populations in microcosm communities at environmentally relevant concentrations. Environ Toxicol Chem 30: 1689-1696.

Belloni V., F. Dessi-Fulgheri, M. Zaccaroni, C.E. Di, A.G. De, E. Testai, et al. 2011. Early exposure to low doses of atrazine affects behavior in juvenile and adult CD1 mice. Toxicology 279: 19-26.

Bishop C.A., S.L. Ashpole, A.M. Edwards, A.G. van, and J.E. Elliott. 2010. Hatching success and pesticide exposures in amphibians living in agricultural habitats of the South Okanagan Valley, British Columbia, Canada (2004-2006). Environ Toxicol Chem 29: 1593-1603.

C.B. Choung, R.V. Hyne, R.M. Mann, N.M. Stevens, and G.C. Hose. 2011. Developmental toxicity of two common corn pesticides to the endangered southern bell frog (*Litoria raniformis*). Environ Pollut 159: 2648-2655.

Collins J.P. and M.L. Crump. 2010. Extinction in our times: Global amphibian decline. Oxford University Press, New York, New York.

Corvi M.M., K.A. Stanley, T.S. Peterson, M.L. Kent, S.W. Feist, J.K. La Du, et al. 2012. Investigating the impact of chronic atrazine exposure on sexual development in zebrafish. Birth Defects Res B Dev Reprod Toxicol.

Cragin L.A., J.S. Kesner, A.M. Bachand, D.B. Barr, J.W. Meadows, E.F. Krieg, et al. 2011. Menstrual cycle characteristics and reproductive hormone levels in women exposed to atrazine in drinking water. Environ Res 111: 1293-1301.

Davis M.J., P. Kleinhenz, and M.D. Boone. 2011. Juvenile green frog (*Rana clamitans*) predatory ability not affected by exposure to carbaryl at different times during larval development. Environ Toxicol Chem 30: 1618-1620.

Fuentes L., L.J. Moore, J.H. Rodgers Jr., W.W. Bowerman, G.K. Yarrow, and W.Y. Chao. 2011. Comparative toxicity of two glyphosate formulations (original formulation of Roundup(R) and Roundup WeatherMAX(R)) to six North American larval anurans. Environ Toxicol Chem 30: 2756-2761.

Hosmer A., W. Kloas, I. Lutz, T. Springer, J. Wolf, and L. Holden. 2007. Atrazine. Response of larval *Xenopus laevis* to atrazine exposure: Assessment of metamorphosis and gonadal morphology. Final Report. Conducted by the Leibniz Institute of Freshwater Biology and Inland Fisheries (IGB), Wildlife International, Ltd., and Experimental Pathology laboratories, Inc . Sponsor: Syngenta Crop Protection. Unpublished. (MRID 471535-01)

Nascimento C.R., M.M. Souza, and C.B. Martinez. 2012. Copper and the herbicide atrazine impair the stress response of the freshwater fish *Prochilodus lineatus*. Comp Biochem Physiol C Toxicol Pharmacol 155: 456-461.

Paetow L.J., J.D. McLaughlin, R.I. Cue, B.D. Pauli, and D. J. Marcogliese. 2012. Effects of herbicides and the chytrid fungus *Batrachochytrium dendrobatidis* on the health of post-metamorphic northern leopard frogs (*Lithobates pipiens*). Ecotoxicol Environ Saf 80: 372-380.

Paulino M.G., M.M. Sakuragui, and M.N. Fernandes. 2012a. Effects of atrazine on the gill cells and ionic balance in a neotropical fish, *Prochilodus lineatus*. Chemosphere 86: 1-7.

Paulino M.G., N.E. Souza, and M.N. Fernandes. 2012b. Subchronic exposure to atrazine induces biochemical and histopathological changes in the gills of a Neotropical freshwater fish, Prochilodus lineatus. Ecotoxicol Environ Saf 80: 6-13.

Shenoy K. 2012. Environmentally realistic exposure to the herbicide atrazine alters some sexually selected traits in male guppies. PLoS One 7: e30611.

Scott C.S. and J. Jennifer. 2011. Trichloroethylene and cancer: Systematic and quantitative review of epidemiologic evidence for identifying hazards. Int. J. Environ. Res. Public Health, 8, pp4238-4272; doi: 10.3390/ijerph8114238 (Open Access at http://www.mdpi.com/1660-4601/8/11/4238/)

Sowers A.D., M.A. Mills, S.J. Klaine. 2009. The developmental effects of a municipal wastewater effluent on the northern leopard frog, *Rana pipiens*. Aquat. Toxicol. 94: 145-152.

Tillitt D.E., D.M. Papoulias, J.J. Whyte, and C.A. Richter . 2010. Atrazine reduces reproduction in fathead minnow (*Pimephales promelas*). Aquat. Toxicol. 99: 149-159.

Topic B References

Abou-Waly H., M.M. Abou-Setta, H.N. Nigg, and L.L. Mallory. 1991. Growth response of freshwater algae, *Anabaena flos-aquae* and *Selenastrum capricornutum* to atrazine and hexazinone herbicides. Bull. Environ. Contam. Toxicol. 46:223-229.

Brain R.A., J. Hoberg, A.J. Hosmer, and S.B. Wall. 2012. Influence of light intensity on the toxicity of atrazine to the submerged freshwater macrophyte *Elodea canadensis.* Ecotoxicology and Environ. Safety, 79: 55-61

deNoyelles, Jr., F., W.D. Kettle, and D.E. Sinn. 1982. The responses of plankton communities in experimental ponds to atrazine, the most heavily used pesticide in the United States. Ecology. 63 (5):1285-1293 (MRID 47543607).

deNoyelles, Jr., F., W.D. Kettle, C.H. Fromm, M.F. Moffett, and S.L. Dewey. 1989. Use of experimental ponds to assess the effects of a pesticide on the aquatic environment. department of systematics and ecology, University of Kansas. Entomological Society of America. Miscellaneous Publications No. 75: 41 – 56 (MRID 47543608)

Fairchild, J.F., T.W. La Point, and T. Schwartz. 1994. Effects of an herbicide and insecticide mixture in experimental aquatic mesocosms. Arch. Environ. Toxicol. Chem. 27:527-533

Fairchild, J.F. 2012. Structural and functional effects of herbicides on non-target organisms in aquatic ecosystems with an emphasis on atrazine. Pp. 383-404, *In*: A. Kortekamp, Ed., Herbicides and Environment, InTech Publishing, ISBN: 978-953-307-476-4. 746 pp.

Giddings J.M. 2012. Atrazine: Review of mesocosm and microcosm studies. Compliance Services International (CSI). May 2012. Sponsor: Syngenta Crop Protection. Unpublished. Available in the public docket, EPA-HQ-OPP-2012-0230.

Herman D., N.K. Kaushik, and K.R. Solomon. 1986. Impact of atrazine on periphyton in freshwater enclosures and some ecological consequences. Can. J. Fish. Aquat. Sci. 43:1917-1925

Hamilton P.B., G.S. Jackson, N.K. Kaushik, K.R. Solomon, and G.L. Stephenson. (1988). The impact of two applications of atrazine on the plankton communities of in situ enclosures. Aquatic Toxicology. 13:123-140.

Hughes J.S., M.M. Alexander, and K. Balu. 1988. An evaluation of appropriate expressions of toxicity in aquatic plant bioassays as demonstrated by the effects of atrazine on algae and duckweed. In: Adams W, Chapman GA, Landis WG (eds). Aquatic Toxicology and Hazard Assessment: 10th Volume. Philadelphia, PA: American Society for Testing and Materials.

Jones D.K., J.I. Hammond, and R.A. Relyea. 2009. Very highly toxic effects of endosulfan across nine species of tadpoles: lag effects and family-level sensitivity. Environ Toxicol Chem 28: 1939-1945.

Relyea R.A. and D.K. Jones. 2009. The toxicity of Roundup Original Max to 13 species of larval amphibians. Environ Toxicol Chem 28: 2004-2008.

McGregor, E.B, K.R. Solomon, M.L. Hanson. 2008. Effects of Planting System Design on the Toxicological Sensitivity of *Myriophyllum spicatum* and *Elodea canadensis* to Atrazine. Chemosphere 73:249-260

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267-1276.

U.S. EPA. 2005. Water quality standards: Tiered aquatic life uses. EPA-822-R-05-001.

Topic C References

Baker D.B., R.P. Richards, T. Timothy, T.T. Loftus and J.W. Kramer. 2004. A new flashiness index: characteristics and applications to Midwestern rivers and streams. Journal of American Water Resources Assoc., 40:503-522

De'ath G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83: 1105-1117. http://dx.doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0CO;2

Lerch R.N., E.J. Sadler, K.A. Sudduth, C. Baffaut, and N.R. Kitchen. 2011. Herbicide transport in Goodwater Creek Experimental Watershed: I. Long-term research on atrazine. J. American Water Resour. Assoc., 47: 209-223.

Miller P.S., W.J. Northcott, J. Malia Andrus, C.M. Harbourt and P. Hendley, 2012. PRZM-Hybrid Modeling system Utilizing and Agronomic Approach to Define Watershed Scale Chemical Applications. (pre-publication draft submitted to the FIFRA-SAP 2012 Atrazine meeting docket).

Olden, J.D. and N.L. Poff. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19: 101-121.

Pappas A. and C. Huang. 2008. Predicting atrazine levels in water utility intake water for MCL compliance. Environ. Sci Technol. 42: 7064-7068.

Prasad A., L. Iverson and A. Liaw. 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9:181-199.

Rabiet M., C. Margoum, V. Guoy, N. Carluer, and M. Coquery. 2010. Assessing pesticide concentrations and fluxes in the stream of a small vineyard catchment – Effect of sampling frequency. Environ. Pollution, 158: 737-748

Temnerud J. and K. Bishop. 2005. Spatial variation of streamwater chemistry in two Swedish boreal catchments: implications for environmental assessment. Environ. Sci. Technol. 39: 1463–1469.

Wagener T. and H.S. Wheater. 2006. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. J. Hydrology 320: 132-154.

Yadav M., T. Wagener, and H. Gupta. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv. Water Resources 30: 1756-1774.

Zanardo, S., N. B. Basu, G. Botter, A. Rinaldo, and P.S.C. Rao. 2012. Dominant controls on pesticide transport from tile to catchment scale: Lessons from a minimalist model. Water Resour. Res. 48, WB04525.

Zhang Z., T. Wagener, P. Reed, and R. Bhushan. 2008. Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, Water Resour. Res., 44: W00B04

FIFRA SAP Meetings

Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) 2003. Report of the FIFRA Scientific Advisory Panel: A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding Potential Developmental Effects of Atrazine on Amphibians. Prepared for June 17 - 20, 2003 meeting of the FIFRA Scientific Advisory Panel. Report No. 2003-01. Document available at EPA-HQ-OPP-2003-0186.

FIFRA SAP. 2007. Report of the FIFRA Scientific Advisory Panel: A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: The Potential for Atrazine to Affect Amphibian Gonadal Development. Prepared for the October 9 - 12, 2007 meeting of the FIFRA Scientific Advisory Panel. Document available at EPA-HQ-OPP-2007-0498

FIFRA SAP. 2009. Report of the FIFRA Scientific Advisory Panel: A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding The Ecological Significance of Atrazine Effects on Primary Producers in Surface Water Streams in the Corn and Sorghum Growing Region of the United States (Part II). Prepared for May 12 – 15, 2009 meeting of the FIFRA Scientific Advisory Panel. Document available at EPA-HQ-OPP-2009-0104.

FIFRA SAP 2011. Report of the FIFRA Scientific Advisory Panel: A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Re- Evaluation of Human Health Effects of Atrazine: Review of Cancer Epidemiology, Non-cancer Experimental Animal and *In vitro* Studies and Drinking Water Monitoring Frequency. Prepared for the July 26-29, 2011 meeting of the FIFRA Scientific Advisory Panel. Document available at EPA-HQ-OPP-2011-0399.