SAP Report No. 2001-08

FIFRA Scientific Advisory Panel Meeting, June 27-29, 2001, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Review of Non-Target Plant Toxicity Tests Under the North American Free Trade Agreement (NAFTA)

SAP Report No. 2001-08

FIFRA Scientific Advisory Panel Meeting, June 27-29, 2001, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Review of Non-Target Plant toxicity Tests Under the North American Free Trade Agreement (NAFTA)

Mr. Paul Lewis
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: January 29, 2002

Fumio Matsumura, Ph.D.
Session Co-Chair
FIFRA Scientific Advisory Panel
Date: January 28, 2002

Ms. Olga Odiott
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: January 29, 2002

Stephen M. Roberts, Ph.D.
Session Co-Chair
FIFRA Scientific Advisory Panel
Date: January 29, 2002

NOTICE

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

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

CONTENTS

PARTICIPANTS ---------- 1

PUBLIC COMMENTERS ---------- 2

INTRODUCTION ---------- 3

CHARGE ---------- 4

SUMMARY OF PANEL RECOMMENDATIONS ---------- 11

DETAILED RESPONSE TO THE CHARGE ---------- 14

LITERATURE CITED---------- 52

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
June 27-29, 2001

Review of Non-Target Plant Toxicity Tests Under the North American Free Trade Agreement (NAFTA)

PARTICIPANTS

FIFRA SAP Co-Chairs
Fumio Matsumura, Ph.D., Professor, Institute of Toxicology and Environmental Health University of California at Davis, Davis, CA

Stephen Roberts, Ph.D., Director, Center for Environmental and Human Toxicology, University of Florida, Gainesville, FL

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

FQPA Science Review Board Members
Celine Boutin, Ph.D., National Wildlife Research Centre, Canadian Wildlife Service, Environment Canada, Hull, Quebec, Canada

Harold D. Coble, Ph.D., Agronomist, USDA/ARS/OA, Office of Pest Management Policy Raleigh, NC

Mr. James F. Fairchild, U.S. Geological Survey, Columbia, MO

Anne Fairbrother, DVM, Ph.D., Senior Wildlife Ecotoxicologist, Parametrix, Inc.,Corvallis, OR

Robert J. Fellows, Ph.D. Senior Research Scientist II, Pacific Northwest National Laboratory Environmental Characterization and Risk Assessment Group, Richland, WA

David Hall, Ph.D., President, David W. Hall Consultant, Inc., Gainesville, FL

Stephen J. Klaine, Ph.D., Professor of Environmental Toxicology, Clemson University, Pendleton, SC

Bruce Maxwell, Ph.D., Land Resources & Environmental Science Dept., Montana State University, Bozeman, MT

Michael Newman, Ph.D., Office of Dean of Graduate Studies, Virginia Institute of Marine Sciences, Gloucester Point, VA

Gladys Stephenson, ESG International Inc., Guelph, Ontario, Canada

Arthur J. Stewart, Ph.D., Science Leader, Ecological Management Science and Technology Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

George E. Taylor, Jr., Professor, Honors Program in General Education and Program in Environmental Science and Public Policy, George Mason University, Fairfax, VA

PUBLIC COMMENT\ERS

Oral statements were made by:

Mr. Tom Gilding representing the American Crop Protection Association
Mr. Michael Leggett representing Crop Protection Institute of Canada
Ms. Jennifer Shaw representing behalf of Syngenta
Ms. Joy Honegger representing behalf of Monsanto
Mr. Michael Dobbs representing Aventis Crop Protection
Mr. Bob McKelvey representing Du Pont
Mr. John Wright representing Dow AgroSciences
Mr. Larry Kapustka representing Ecological Planning and Toxicology, Inc.
Kassim Al-Khatib, Ph.D., representing Kansas State University
Kurt Getsinger, Ph.D. representing the US Army Engineer Research and Development Center
Jeffrey Giddings, Ph.D., representing the Cadmus Group, Inc.
Diana Post, Ph. D., representing the Rachel Carson Council, Inc.

Written statements were made by:

None were received.

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency pertaining to its review of non-target plant toxicity tests under the North Atlantic Free Trade Agreement (NAFTA). Advance notice of the meeting was published in the *Federal Register* on May 4, 2001. The review was conducted in an open Panel meeting held in Arlington, Virginia, June 27-29, 2001. The meeting was co-chaired by Fumio Matsumura, Ph.D. and Stephen M. Roberts, Ph.D. Mr. Paul Lewis and Ms. Olga Odiott served as the Designated Federal Officials. The issues to be discussed were presented jointly by the EPA Office of Pesticide Programs (OPP) and by Health Canada's Pest Management Regulatory Agency.

The SAP was asked to provide recommendations on proposed testing levels, test species, test protocols, endpoints, triggers to move between testing levels, potential use of uncertainty factors, and research needed to improve plant testing.

CHARGE

Question 1 Aquatic and terrestrial "Level" assessment scheme

Four assessment levels are proposed for determining chemical risk to non-target plants. The levels progress from a deterministic assessment (Level 1) to more refined assessments (Levels 2-4). Groups of species identified as being at risk at one level will trigger assessment at higher levels. There are various options for refining aquatic and terrestrial plant assessments. These include a more realistic exposure estimation, additional species testing (to improve characterization of sensitivity), or examination of the ratio of an exposure distribution over a sensitivity distribution (rather than a ratio of single points). These refinements are intended to reduce the level of uncertainty in an assessment.

What are the advantages and disadvantages of the proposed tiered approach for assessing chemical risk to plants?

What refinements to plant assessment are recommended at each level?

What are the Panel's suggestions for an improved or more detailed plant assessment scheme beyond that already discussed?

Question 2 Uncertainty factors

Uncertainty factors can be used in estimating the risk of chemicals to aquatic and terrestrial plants. Uncertainty factors have been defined in a number of ways, but most are related to two concepts: 1) addressing uncertainty due to variability in testing and extrapolation to untested species, and 2) addressing uncertainty due to the limited data. The agencies have proposed expansion of test species to address these issues; however, uncertainty factors may also improve confidence in risk assessments.

What are the advantages and disadvantages of using uncertainty factors in lieu of additional aquatic or terrestrial plant toxicity tests? How should they be applied?

What are the Panel's thoughts on the applicability of aquatic plant uncertainty factors (as currently used by the OPPT) to terrestrial plants? Are entirely different uncertainty factors needed?

Question 3 Aquatic triggers for progression to higher levels

Effective concentration (EC) values are used to determine chemical risk to aquatic plants. Population endpoints (e.g., biomass as measured in algae) are measured over a full life cycle, whereas individual endpoints (e.g., plant length as measured in *Myriophyllum* spp.) can be measured partially or over a full life cycle. Various options have been presented for selection of EC values.

What aquatic EC values (NOAEC, EC_{50}, or other) are appropriate for this testing scheme

and for calculating risk quotients at Level 1?

What aquatic EC values (e.g., EC_{10}, EC_{25}, etc.) should be used for population and individual plant parameters?

Question 4 Terrestrial triggers for progression to higher levels

Progression through the level system is influenced by the sensitivity of the plant and exposure to the test chemical. Detection of this sensitivity depends on the endpoints selected, as well as when they are measured during the plant life cycle. In current assessments, individual plant parameters (e.g., phytotoxicity, height, dry weight) are measured for the determination of chemical risk to non-target terrestrial plants. Currently, population effects are not predicted but the EPA and PMRA expect to address these impacts in the near future. At the present time, the EC_{25} value for the most sensitive parameter is used in the calculation of a risk quotient at Level 1.

What parameters (height, dry weight, survival, etc.) are appropriate to measure at Level 1? Which EC value should be used for each parameter?

What are the advantages and disadvantages of each parameter with respect to the expense and time involved in conducting the test and the ability of the test results to accurately predict effects?

How can individual plant parameters be used to predict population effects?

Question 5 Consideration of physiological and biochemical markers

In addition to established measurement endpoints (gross acute), recent literature suggests it is important to consider alternative endpoints pertaining to physiological and biochemical effects (e.g., O_2 production, carbon fixation, etc.). The endpoints assessed in seedling emergence and vegetative vigor tests may not fully detect physiological changes that are detrimental to plants.

How important is it to consider physiological or biochemical endpoints when determining chemical risk to plants?

What physiological or biochemical changes would be considered adverse? Please quantify (e.g., 20%, 50%, etc.).

What would be the advantages and disadvantages of replacing seedling emergence and vegetative vigor endpoints with physiological and biochemical endpoints?

Which additional endpoints could be used to assess hazard to non-target plants?

Question 6 Aquatic and terrestrial reproductive effects

Because existing plant toxicity tests do not provide an adequate trigger for reproductive

testing, the PMRA and the EPA propose that reproductive testing be conducted at Level 1. For aquatic plant assessments, we propose the testing of two species (rice and nodding smartweed). For terrestrial plant assessments, we propose two of three test species [cherry, mouse-ear cress (*Arabidopsis thaliana*), and canola (*Brassica rapa*)]. The agencies request guidance on how to refine the assessment of reproductively sensitive plant species.

What are the advantages and disadvantages of the proposal to assess reproductive effects at Level 1 versus higher levels?
What are the advantages and disadvantages of partial- versus full-life cycle testing?

For full life-cycle testing, what are the advantages and disadvantages of testing mouse-ear cress versus canola?

What are some alternative approaches and species that address potential reproductive effects?

What reproductive testing endpoints (e.g., pollen viability, seed formation, etc.) are critical for Level progression and for assessing chemical risk to non-target plants?

What reproductive EC values (e.g., NOAEC, EC_{05}) are critical for Level progression and for assessing the risk to non-target plants?

How critical is it to consider both modes of reproduction (sexual versus asexual)?

Question 7 Marine algal toxicity testing

The range of response among different marine algae can be as great as is observed in freshwater species. Currently, only one marine algal species test is required at Level 1. Researchers have recommended a test battery which includes species of diatoms, green algae, and dinoflagellates as well as golden-brown algae. This test battery should provide a range of responses until sufficient comparative toxicity studies are available to determine sensitivity ratios. In addition to the currently required marine diatom, the proposed plant testing scheme requires the testing of one marine algal species in each of three previously unrepresented Divisions (Chyrsophycophyta, Pyrrhophycophyta, and Rhodophyta).

Are the four proposed marine algal species representative? What other marine algal species should be considered? Why?

What are the limitations of the protocols and the availability of the proposed marine algal species?

Question 8 Aerial exposure testing of floating aquatic plants

Chemical exposure can occur from drift deposited onto the leaf surfaces of floating

plants. For contact toxicants, it is not sufficient to test only aquatic exposure (i.e., chemical dissolved in water). It has been observed that the sensitivity of *Lemna* to a contact herbicide can increase several-fold with a foliar exposure compared to the conventional exposure through the growth medium. Although there are limited data on aerial exposure, both Agencies believe that this type of study should be conducted on a routine basis.

What are the pros and cons of routinely requiring foliar exposure tests for floating aquatic plants?

How can the methodology by Lockhart *et al.* (1989) be modified for future testing requirements?

What research needs can be identified for foliar exposure testing?

Question 9 Submersed aquatic vascular plant testing

Submersed aquatic vascular plants are morphologically different compared to terrestrial vascular plants or algae. Currently, *Lemna* spp. (floating) are used to predict effects on submersed aquatic vascular plants. As *Lemn*a is a monocotyledon, it would be preferable to also require a representative dicotyledon. The agencies recommend *Myriophyllum* spp. to represent dicotyledonous submersed aquatic species.

What is the SAP's opinion on the proposed toxicity test requirement with a submersed aquatic vascular species?

What are the Panel's thoughts on the selection of *Myriophyllum* spp. to represent dicotyledonous submersed aquatic vascular plants?

What other monocotyledonous and dicotyledonous submersed species should be considered? Why?

Question 10 Emergent aquatic vascular plant testing

Currently, *Lemna* spp. (floating) and terrestrial vascular plants are used to predict the effects of toxicants on emergent aquatic vascular plants. However, there are no available data to support the use of these as surrogates for emergent aquatic plants. Many emergent aquatic plant species prosper in both aquatic and terrestrial habitats. In an aquatic exposure scenario, emergents are unique in that they can be exposed *via* root or stem uptake of contaminated water in addition to foliar exposure from an over-spray.

What are the Panel’s thoughts regarding the physiological differences between emergent aquatic and terrestrial vascular species or *Lemna* sp. to support the proposed approach?

What are the Panel’s thoughts on the selection of rice and nodding smartweed to represent emergent aquatic vascular plants?

What other emergent aquatic vascular plant species can the Panel recommend to satisfy both terrestrial and aquatic testing requirements?
How important is it to consider routes of exposure other than foliar exposure to emergent aquatic species (i.e., absorption from the water column by the submersed stem or from sediments *via* the roots)?

What is the Panel's opinion regarding extrapolation from terrestrial vascular species or *Lemna* spp. to emergent rooted aquatic vascular plants?

Question 11 Mode of action

The agencies are proposing a seedling emergence and a vegetative vigor test at Level 1. Various factors contribute to eliciting toxicity of a particular chemical (e.g., mechanism of uptake by a plant, the mode of action of the chemical within a plant, application parameters, etc.). Information on these factors could reduce the need to conduct a seedling emergence or vegetative vigor test.

What criteria might we use for data waivers based on the mode of uptake and the mode of action of a chemical?

How could the number and types of tests be reduced with respect to application parameters (i.e., timing, method, etc.)?

What other factors might reduce the amount of testing required?

Question 12 Terrestrial Species

Both Agencies have recognized the need to consider sensitivity to chemicals among a broad range of ecologically-relevant plant families. The Agencies have proposed to increase the number of families tested to reduce uncertainty and variability with respect to sensitivity. Researchers have recommend a test battery including non-crop and woody species to encompass a range of response until sufficient comparative toxicity studies are available to determine sensitivity ratios. The selection of the families and the species within those families was based on the feasibility of using them as test species and their economic or ecological importance.

What are the Panel's thoughts on the proposed terrestrial species at Level 1?

What (if any) additional species or groups are not adequately represented in the proposed testing scheme at Level 1?

Are there better approaches for selection of species besides the taxonomic / phylogenetic approach (e.g., ecological or functional approach)? What are they?

How can the agencies improve their knowledge of the variability in sensitivity of the proposed test species?

Question 13 Additional Species Testing

Level 2 is envisioned as primarily an assessment level that utilizes refined exposure methods and toxicity assessment. However, additional testing may be needed to clarify uncertainties before advancing to Level 3, such as laboratory to field extrapolation or specific dose-response curves. Level 3 is envisioned to include expansion of testing in two areas: reproductive testing and acute testing of keystone species. Keystone species can be selected in a couple of ways: (1) keystone species **within families** triggered by risk identified at Level 2; and (2) keystone species within new families that are **within a structure/function group** (e.g., woody plants) identified to be at risk at Level 2 or identified in incident reports.

What are the Panel's thoughts on additional testing to clarify uncertainties on previously tested species in Level 2?

What are the Panel's thoughts on having two areas of focus in Level 3 (reproductive testing and acute testing of keystone species)?

What are the advantages and disadvantages of additional species testing at Level 3? Should the additional species be focused on keystone or ecologically significant species prevalent in areas of chemical use?

What are the Panel's thoughts on expanded testing of species in sensitive structure/function groups?

Question 14 Aquatic and terrestrial multi-species testing

Multi-species testing is proposed at the comprehensive level of assessment (Level 4). Population dynamics and community structure could be affected due to differences in chemical sensitivity among individual species. This may result in an alteration of plant community structure which subsequently may lead to adverse effects on organisms at higher trophic levels. Multi-species studies provide necessary and invaluable information about changes in population and community dynamics that result from phytotoxic impacts.

How useful are data generated from multi-species/community level studies?

When is multi-species testing appropriate in the proposed design (i.e., how should it be triggered)?

How many trophic levels should be considered in a multi-species test when considering the risk of chemicals to plants?

Question 15 Aquatic and terrestrial post-registration monitoring

Post-registration monitoring is proposed at the comprehensive level of assessment (Level 4) when adverse effects are anticipated for sensitive species or groups (identified at Level

3). The location and number of monitoring studies will depend on the sensitive species or groups identified and on the types of eco-regions in which they occur. A monitoring study can focus on an indicator species expected to be sensitive, or a multi-species testing design can be introduced to consider the effects on the whole community.

What are the advantages and disadvantages of monitoring studies focused on indicators versus multi-species (communities)?

What criteria are used in the selection of an indicator plant species?

Question 16 Bioaccumulation

There is a potential for bioaccumulation of chemicals in non-target plants. Bioaccumulation may be one indicator of hazard and more importantly an indication of the extent of uptake and translocation of chemicals in plants. Chemicals that bioaccumulate in plants may also have implications for herbivorous wildlife species. The Agencies are less certain on whether to assess the effects of bioaccumulation in the determination of overall risk to non-target plants.

What are the SAP's thoughts on the need for uptake / accumulation tests to address bioaccumulation in plants?

How should the agencies address bioaccumulation?

Question 17 Research

Since the last SAP meeting, ORD has developed test methods, including *Lemna* and *Arabadopsis* life cycle tests. ORD has also conducted comparative toxicity laboratory and field studies for herbicide effects on annual and woody plants. In addition, ORD has studied short- and long-range transport of chemicals, such as ozone and acid rain, and potential impacts of their deposition on sensitive plants, including endangered and forestry species. A long-range transport model was developed by EPA/Duluth and has been used to model atrazine herbicide transport.

What are the most important short-term (5 years) and long-term (10 years) research initiatives that will improve plant toxicity testing for the regulation of chemicals?

SUMMARY OF PANEL RECOMMENDATIONS

Key conclusions and recommendations from the Panel are listed below:

The Panel did not agree with the aquatic and terrestrial assessment scheme as proposed. Rather than move to a new, expanded assessment scheme, the Panel suggested a more methodical approach that begins with problem formulation, incorporates a tiered approach for plant testing, and reduces uncertainty in the decision-making process with each successive level. The Panel supports a limited number of test species in Level 1 in association with a highly conservative exposure estimate. If this trigger is exceeded, then an addition of more test species, along with a refined exposure estimate using the GENEEC or other model is warranted. The process should be incorporated with a problem formulation context that considers cropping patterns, application methods, and other critical components of the risk assessment process.

The Panel strongly recommended that an Agency-Industry-Academia group be organized to consider standard methods of collecting and scoring efficacy data compiled during the initial screening of a new chemical to maximize their use in problem formulation.

The Panel concluded that the use of conservative exposure scenarios coupled with a conservative response matrix obviates the need for uncertainty factors for Level 1. As one proceeds through the Levels, the use of species sensitivity distributions, rather than single point estimates of the EC_{25} from the most sensitive species, provides the required level of certainty that the hazard assessment is applicable to all plant taxa. Rather than addressing the potential use of uncertainty factors, EPA should focus efforts on determining which species provide the most robust description of species sensitivity distributions.

Concerning aquatic triggers for the testing scheme and risk quotients at Level 1, the Panel agreed that progression from IA to IB should be based on a statistically significant effect >10% relative to the control. It was also stated that progression from 1B to II should be based on EC50/10 for population growth parameters and EC25/10 for individual growth parameters.

The Panel agreed that biomass, because it is a structural endpoint, should be measured at Level 1. Most members of the Panel agreed that other endpoints such as height, visual assessment, and root weight, should be used depending on the species, the type of application, type of chemicals, etc.

The Panel cautioned that extrapolation from individual to population endpoints cannot be done with confidence, particularly when many of the endpoints lack a basis in functional population biology. Any transfer from individual to population endpoints would require a modeling approach that factors in species-specific attributes.

Prior to replacing seedling emergence and vegetative vigor endpoints with physiological

and biochemical endpoints, the Agency should convene a Work Group of biomarker experts to discuss the state-of-the art, review various protocols, and consider the applicability of such endpoints to ecological and agronomic endpoints.

The Panel suggested the following endpoints could be used to assess hazards to non-target plants:

- Whole plant physiology including net photosynthesis, stomatal conductance, photosynthate allocation
- Remote sensing using HSI technology
- Genomics
- Community Importance index (CI)

Reproductive testing should not be required at Level 1. Instead, this should be an option available if a concern is identified during problem formulation. If the mode of action suggests that reproductive effects could occur, this test should be incorporated into Level 1. Research to better understand if reproductive effects occur in the absence of other, currently measured effects should be strongly supported.

Level 2 should provide better characterization of exposure scenarios. Once the potential regional use patterns are established and specific areas vulnerable to exposure are identified for use of the pesticide or chemical, a list of test species for those regions or areas could be identified.

Level 3 should utilize strategic additional species testing to develop a distribution of sensitivities within plant groups that have been shown to be sensitive to the herbicide of concern. The focus should be on sensitive species that are present in the area of intended use, and the studies must be designed to answer very specific questions. This would facilitate a probabilistic risk assessment.

Rather than expanding the number of species tested, the Agency should first determine the optimal number of species required to define the species sensitivity distribution and then determine the optimal species representation. Thus, the extant data base should provide an empirical basis for this determination.

The Panel noted that not all incidences of non-target adverse ecological effects are the result of off-site migration of pest control products from the site of application. It is possible that some of these incidences could have been caused by non-chemical stressors. Therefore, monitoring programs that are invoked post-registration must be able to discern between chemical (e.g. herbicide) and non-chemical (e.g. water stress; insect damage, disease, etc.).

The fate of chemicals in an ecosystem should be known with some degree of certainty, and one of the endpoints of concern is bioaccumulation. Bioaccumulation is important because of the potential for food chain transport and biomagnification. Most Panel members agreed

that chemicals that show bioaccumulation should be investigated more fully than their counterparts that tend to be more uniformly distributed.

At the request of the Agency, the Panel provided a list of short-term (5 years) and long-term (10 years) research initiatives for the Agency to consider. The Panel emphasized that the risk assessment framework needs to be developed and applied to pesticide registration. The Agency should develop the approach for deriving the problem formulation portion of each risk assessment and Registrants should be required to submit this to the Agency along with Level 1 data. The Panel also stated that it appears absolutely crucial that a stakeholders meeting be held to discuss standardization and research needs.

DETAILED RESPONSES TO THE CHARGE

Question 1 **<u>Aquatic and terrestrial "Level" assessment scheme</u>**

Four assessment levels are proposed for determining chemical risk to non-target plants. The levels progress from a deterministic assessment (Level 1) to more refined assessments (Levels 2-4). Groups of species identified as being at risk at one level will trigger assessment at higher levels. There are various options for refining aquatic and terrestrial plant assessments. These include a more realistic exposure estimation, additional species testing (to improve characterization of sensitivity), or examination of the ratio of an exposure distribution over a sensitivity distribution (rather than a ratio of single points). These refinements are intended to reduce the level of uncertainty in an assessment.

What are the advantages and disadvantages of the proposed tiered approach for assessing chemical risk to plants? What refinements to plant assessment are recommended at each level? What are the Panel's suggestions for an improved or more detailed plant assessment scheme beyond that already discussed?

The Panel supports the use of a tiered assessment process because it allows optimum use of financial and intellectual resources in ensuring environmental protection. Any assessment scheme must be based on a clear understanding of the problem to be addressed. The proposed assessment scheme increases the number of test plant species from 10 to 26 and increases the numbers of both monocotyledon and dicotyledon families. It is not apparent, specifically for Level 1, that this increase will result in better protection of non-target species. The Panel endorses an approach that parallels the ecological risk paradigm. In addition, much information can be mined from the efficacy data compiled during initial screening of the new chemical. It is critical that an Agency-Industry-Academia work group be organized to standardize the methods of collecting and scoring these data to maximize their use in problem formulation.

As proposed, the assessment scheme has little scientific foundation and appears to be driven by a perception that impacts to non-target plants are occurring at an unacceptable rate. Rather than move to a new, expanded assessment scheme, the Panel suggests a more methodical approach that begins with problem formulation, incorporates a tiered approach for plant testing, and reduces uncertainty in the decision-making process with each successive tier. An assessment scheme would begin with a problem formulation phase that would identify important exposure and effects issues. A clear and relevant conceptual model is generated around which the assessment is to be constructed and then the most appropriate information is gathered together. This would result in the selection of test species based on likely exposure scenarios and an understanding of the mode of action of the new chemical. In such a scheme, it would be best to have a battery of test plant species with standardized testing procedures for each. Initial testing should be conservative in both exposure scenario and measured effects. In addition, Level 1 should provide data upon which clear decisions can be made about moving the chemical testing to the next tier. The use of the risk quotient for this Level is appropriate.

In Level 2, the expanded focus on better characterization of the exposure scenarios will be beneficial and should decrease the uncertainty raised in Level 1.

In Level 3, expansion of the test species to enable a probabilistic risk assessment is an excellent approach. Additional test species should be chosen based on results of Level 1 and a better understanding of ecosystems at risk. The focus on an improved understanding of the exposure scenarios should permit identification of candidate plant species for testing. In probabilistic risk assessment it is important that attention be paid to determining what exactly is being protected and then formulating a specific set of data requirements. It is important to consider adequate representation of the taxonomic group(s) of concern and required number of species. At present, it is not obvious that the research has been conducted to adequately assess the number of species needed to develop a sufficiently robust species sensitivity distribution. While this is an important research need, one of the most salient and recurrent findings of the science of genomics is the unexpected high degree of homogeneity at the molecular levels between many forms of animals, plants, and microbes. In essence, we share far more in common at the molecular level than differences. Despite this finding, the proposed ecotoxicity testing schemes emphasize supposedly "monumental" differences among species, and these differences are the basis for offering more than 20 species to evaluate ecotoxicity of new chemicals. There is a very strong argument for reducing the number of species to be tested, perhaps relying on more functional attributes as a means of evaluating response. Perhaps the old database for effects of phytotoxic chemicals biases this outcome to reinforce the idea that a large number of species are needed. In essence, it may not be the number of species that is important; rather, it is which species that is critical.

Interpretation of results should be limited to a statement about making a better estimate of a hazardous concentration relative to the risk quotient method. It should not include statements about predicting species population persistence or community integrity unless the metrics used in the probabilistic risk assessment were relevant to those levels of biological organization.

Mortality (number of individuals or biomass loss) data only addresses one facet of population consequences: D in the relationship $dN/dt = rN[1-N/K] = [B-D]N[1-N/K]$ where r = instrinsic rate of increase or the difference between the birth and death rates (B-D), N is the population size, and K is the carrying capacity. Mortality is only one component of this simple equation which describes population change through time.

Stresor effects on crucial species interactions also remain undefined:

where a_{Ij} = interaction coefficients quantifying the influence of other interacting species on the population dynamics of species *i* in a specific system. There may be only a few or many significant species interaction coefficients. Analysis of consequences of species

persistence requires knowledge of community ecology and complex matrix methods including exposure-related changes of all species carrying capacities and competition coefficients. The simple metrics generated at Level 1 are poor predictors of persistence in an exposed field community of nontarget species. As a relevant example, the dynamics and persistence of *Skeletonema costatum* in the spring bloom of Long Island Sound are very dependent on competition with other phytoplankton for nutrients. Competition and other species interactions are often extremely important relative to species population persistence and behavior. Ignoring the role of competition in this example would result in a poor prediction of species density and persistence.

Another suggestion for a community impact metric is the use of the Community Importance index (*CI*) for a set of test species as suggested by Power et al., 1996.

where t_N is a community measure when the community is intact, t_D is the same community measure when species *i* is deleted from the community and *p* is the proportional abundance of species *i* when the community is intact. This endpoint measure could most effectively be used in natural ecosystems or macrocosm experiments. The *CI* could be measured in exposed and unexposed plots for a sensitive species identified at a lower level to quantify community impacts.

In Level 4, it is important to incorporate Geographic Information System (GIS) and remote sensing methodologies in the field testing approach.

Question 2 <u>Uncertainty factors</u>

Uncertainty factors can be used in estimating the risk of chemicals to aquatic and terrestrial plants. Uncertainty factors have been defined in a number of ways, but most are related to two concepts: 1) addressing uncertainty due to variability in testing and extrapolation to untested species, and 2) addressing uncertainty due to the limited data. The agencies have proposed expansion of test species to address these issues; however, uncertainty factors may also improve confidence in risk assessments.

What are the advantages and disadvantages of using uncertainty factors in lieu of additional aquatic or terrestrial plant toxicity tests? How should they be applied?

There are three main sources of uncertainty and variability in a risk assessment: 1) inherent biological/ecological variability, 2) imprecision and inaccuracy of measurements, and 3) insufficient knowledge. Inherent intraspecies biological variability is dealt with in risk assessments through the use of dose-response relationships (after all, if there was no biological

variability, all organisms would respond at the same dose, so we would need to know only the EC_{100}). Impression and inaccuracy of measurements frequently gets folded into the dose-response relationship through the use of uncertainty bounds. However, the most intractable component of uncertainty is our lack of knowledge of the entire process about which we are assessing risk.

In the context of risk assessments conducted for herbicide registration, one is confronted with uncertainty in both the exposure and hazard components. A sensitivity analysis should be used to determine where the greatest uncertainties lie. In ecological risk assessments the greatest uncertainties usually lie in the estimation of exposure rather than the hazard assessment. Nevertheless, this does not negate the fact that we lack certain knowledge of how a new herbicide might affect all aspects of the life cycle of all plants. Therefore, uncertainty in the hazard assessment remains. The incorporation of the mode of action data in the problem formulation phase may reduce this uncertainty.

Uncertainty factors are frequently used in regulatory science when there are limited data; higher uncertainty factors are used for the most limited datasets. The only advantage of using an uncertainty factor is that it ensures that limited data are protective. However, use of uncertainty factors will usually result in highly conservative estimates of risk and therefore, restrictions on chemical use in situations where the actual risk may be quite low. Uncertainty factors are inherently difficult to assign due to the variation both among species responses and the slope of the dose response curve, which typically varies depending on the chemical's mode of action.

Uncertainty factors are used by the Agency and other regulatory organizations when faced with extreme lack of information and the desire to be protective (i.e., to reduce Type II error so as to not declare a chemical safe when it really is not). However, herbicide registration is a very data rich process. Information is available about the mode of action of the chemical and its effects on target species. In addition, several additional plant species (currently 10) are tested for adverse effects as well. Therefore, it is likely that the uncertainty in extrapolating the available information to all nontested species is considerably lower than for other regulatory processes, such as PMNs under TSCA. Many regulatory agencies that use uncertainty factors in setting environmental criteria (e.g., Environment Canada) reduce the uncertainty factor in proportion to the amount of available information. For example, in setting soil criteria, Environment Canada requires that a safety factor of 10 be applied if data for only one plant species is available, a safety factor of 5 if 2 species from different families have been tested, etc. Adopting this philosophy, and considering that the Agency requires information on 10 test species from several families, it should be presumed there is sufficient knowledge to not require the application of any uncertainty factor.

However, in reality, herbicide risk assessors have imperfect knowledge of potential hazard to all species and require reassurance that the risk of making a Type II error is low. In other words, how much information is enough to be convinced that the assessment is sufficiently conservative? If the intent is to find the most sensitive species to test for each chemical, then there will never be enough information and an uncertainty factor approach may become

necessary. However, if the intent is to be reasonably protective of all plants, then the goal should be to adequately describe the distribution of species sensitivities such that a protection level can be determined with reasonable accuracy. Robert McKelvey of DuPont, in his written and oral comments for this SAP meeting, summarizes the work he and others did to generate species sensitivity distributions using historic testing and efficacy data. He noted that six to eight species are sufficient to describe the species sensitivity curve; adding additional species filled in the distribution but did not change its shape. This has been shown to be true for other systems as well. For example, response of aquatic organisms to metal contamination is adequately described by eight species. This provides sufficient information to be protective of aquatic invertebrates (including shellfish), fin fish, and plants – a much wider taxonomic distribution than that being considered in nontarget herbicide assessments. The topic of application of species sensitivity distributions is addressed most notably by the Dutch, including Aldenberg and Slob (1993). A book on this topic is in preparation and should be available by the end of the year (Posthuma and Suter, 2001).

Another uncertainty is how to move from the individual level effects assessment to the population or community level assessment. There was some discussion about using the species sensitivity distribution as a means of protecting community integrity, with varying opinions about how well this approach might work. The most important aspect is to be sure that the species represented in the distribution are pertinent to the ecosystem at hand. Thus, if the assessment is for all of North America, the distribution would need to encompass all species; while if the assessment is only for the corn-belt of Iowa, fewer species would be requuired. Likewise, the endpoint used to establish the sensitivity must be ecologically relevant.

The Panel concluded that the use of conservative exposure scenarios coupled with a conservative response matrix obviates the need for uncertainty factors for Level 1. Use of species sensitivity distributions, rather than single point estimates of the EC_{25} from the most sensitive species, will provide the required level of certainty that the hazard assessment is applicable to all plant taxa. Rather than addressing the potential use of uncertainty factors, EPA should focus their efforts on determining which species provide the most robust description of species sensitivity distributions.

The Panel supports the progression of a tiered approach to reduce the uncertainty due to insufficiency of knowledge through the acquisition of data, rather than the application of uncertainty factors.

What are the Panels thoughts on the applicability of aquatic plant uncertainty factors (as currently used by the OPPT) to terrestrial plants? Are different uncertainty factors needed?

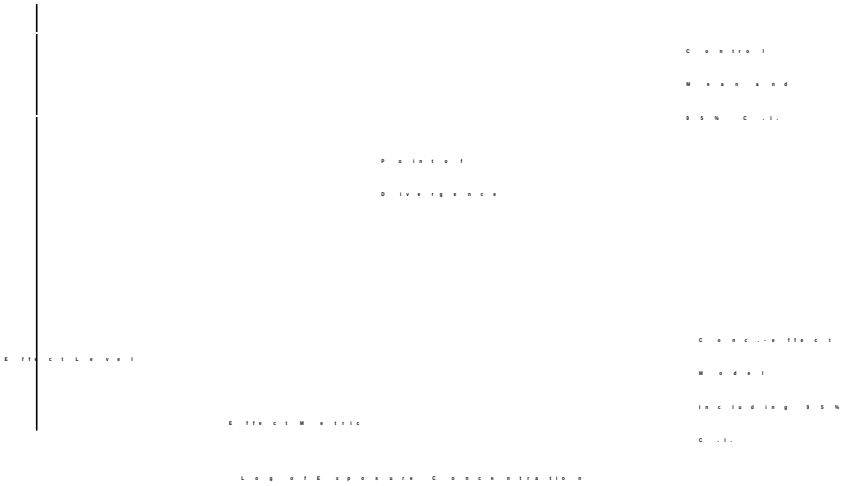
As discussed above, the Panel believes that the use of uncertainty factors should be replaced by a better understanding of which species and how many species need to be tested to adequately represent the species sensitivity distribution. The size of the dataset needed for risk assessment of terrestrial plants is discussed in other questions submitted to the Panel.

Question 3 <u>**Aquatic triggers for progression to higher levels**</u>

Effective concentration (EC) values are used to determine chemical risk to aquatic plants. Population endpoints (e.g., biomass as measured in algae) are measured over a full life cycle, whereas individual endpoints (e.g., plant length as measured in *Myriophyllum* spp.) can be measured partially or over a full life cycle. Various options have been presented for selection of EC values.

What aquatic EC values (NOAEC, EC50, or other) are appropriate for the testing scheme and for calculating risk quotients at Level I?

Progression from IA to IB should be based on Option 2: a statistically significant effect >10% relative to the control. The EPA uses a 50% or greater inhibitory effect level. That may not be sufficient to preclude unreasonable risk. The proposed Canadian guidelines suggest that any statistically significant effect be used as the trigger. This option requires clear control of statistical power in order to be useful, and has the potential risk of spending effort on an effect/level that is not ecologically significant.



Use of hypothesis testing, i.e., NOAEL, is statistically questionable. A regression method to produce a parametric model is preferred. However, if an endangered species is the focus of the assessment, a method that does not imply a certain level of "taking" is preferred. The context of "unreasonable risk" depends on whether one is conducting a general analysis for "non-target species" or an endangered species. A regression method analyzed as per Liber et al.(1992), showed that the effect metric was the

concentration at which the 95% confidence intervals for the control and the concentration-effect lines no longer overlapped. This approach is a relatively common statistical approach found in various introductory biometry textbooks so there is ample justification for its use. Alternatively, the hypothesis testing method could be used if sufficient attention was devoted to the issue of statistical power. Instead of using the LOEC estimated from the highest treatment concentration having a mean effect statistically "different" from the mean of the control, the minimum significant difference statistic could be used. This statistic shows how much of an effect one would have to have in order for it to be seen as "statistically significantly different" from the control mean. Bioequivalence testing methods are also viable, but uncommon, alternatives to ANOVA/post-ANOVA hypothesis testing.

What aquatic EC values (e.g., EC10, EC25, etc,) should be used for population and individual plant parameters?

The Panel agreed that until metrics of population effect are generated in terms that population biologists use to project population fate under various conditions, the answer to this question can only be an informed guess. Depending on the reproductive potential of the species and the level of mortality, the population may or may not disappear. The presence or absence of a substantial propagule or seed bank may also be critical for some species.

The exception would be for risk assessments focused on an endangered species. In these instances, the phrase "unreasonable risk" assumes a different meaning and a regression method applied as per Liber et al.(1992) would be preferred. Functional redundancy of phytoplankton might make the difference between individual-based versus population-based metrics. For atrazine, published studies have shown that an EC50 was fairly predictive of community consequences (community structure and function).

Some Panel members stated that it is difficult to justify changing the current approach without data to illustrate that it is inadequate. The use of an EC50 is already conservative, for it is based on a non-lethal endpoint. In addition, no one has yet refuted it with data.

Question 4 **<u>Terrestrial triggers for progression to higher levels</u>**

Progression through the level system is influenced by the sensitivity of the plant and exposure to the test chemical. Detection of this sensitivity depends on the endpoints selected, as well as when they are measured during the plant life cycle. In current assessments, individual plant parameters (e.g., phytotoxicity, height, dry weight) are measured for the determination of chemical risk to non-target terrestrial plants. Currently, population effects are not predicted but the EPA and PMRA expect to address these impacts in the near future. At the present time, the EC_{25} value for the most sensitive parameter is used in the calculation of a risk quotient at Level 1.

What parameters (height, dry weight, survival, etc.) are appropriate to measure at Level 1?

Which EC value should be used for each parameter?

The use of a battery of endpoints was supported by the Panel although it was argued that the number of endpoints should be kept to a manageable number. There was agreement that biomass should be measured because it is a structural endpoint. Other endpoints such as height, visual assessment, and root weight, should be used depending on the species (e.g. morphology, seed size), the type of application, type of chemicals, etc. The selection might be based on endpoints that have some functional basis in ecology and/or physiology (e.g. carbon allocation, carbon acquisition). Growth analysis, particularly in those instances where the harvesting may be staged over the growing season can determine both chronic (temporary inhibition of growth as dry matter accumulation) and acute (significant growth impairment, death) responses to a stressor. It would be optimal for both plant components to be measured, although root mass may be the more difficult to obtain because of difficulties in removing the planting mixture from the root surface. These tests can be performed readily and relatively inexpensively in an experimental design capable of providing a statistically sound assessment of natural plant-to-plant variation within species or cultivar.

Mode of action of the pesticide tested should be considered when deciding on the endpoints to be measured. There should be some assemblage of responses so that not just one endpoint out of 10 is sufficient to trigger progression to the next Level. By chance alone, one in ten endpoints will show a statistically significant effect at P=0.10. Why not look at all the endpoints measured and use them collectively in a multivariate format? This type of assessment could be used to explain more about potential mode of action on different groups of plants and encompass a wider range of sensitivities. A new generation of investigative tools such as whole-plant physiology, genomics, and remote sensing should be incorporated into the assessment.

In general, the lower the Level, the more conservative the trigger should be to avoid false negatives. The range should be between EC50 and EC25. The issues discussed previously in Question 3 regarding the statistical validity of alternative statistical approaches are pertinent to this discussion.

What are the advantages and disadvantages of each parameter with respect to the expense and time involved in conducting the test and the ability of the test results to accurately predict effects?

Depending on the species some parameters may not be appropriate, e.g. height for grasses. Visual assessment or visible symptomology is a quick and easy endpoint that may or may not be appropriate. The use of seed germination has some potential but it is not a very sophisticated or functionally important parameter simply because approximately 95+% of all seeds either die before germination or shortly thereafter depending on the species. Shoot height may become subjective to measure in those species with prostrate growth or significant branching patterns (e.g., strawberries or soybeans) and can vary between species with light quality and quantity even within the same greenhouse or field plot. Root length measurements can suffer similar problems particularly for species with reduced primary roots or tap roots whose branching patterns are

fibrous. Extraction and separation from the potting medium can be both difficult and time consuming. Seedling emergence, which primarily depends on stored nutrient reserves, may be less influenced by the contaminant in question than by variations in seed quality, seed positioning, pathogenic microorganisms in the rooting mixture, or microclimate effects at the soil surface. At the same time, survival appears to be less critical than some other functional endpoints of a physiological nature. Many of the parameters that are listed as candidates would fall out of contention since they could be easily linked to whole-plant physiology (e.g., peroxidase, fluorescence). Biochemical tests such as enzyme assays, chlorophyll content and/or fluorescence may be more sensitive but suffer the problem of sampling variations within a plant canopy unless the entire plant was "homogenized." As stated in the briefing literature the relationship of these tests to population responses is tentative. Reproductive success could be a potential addition to the dry weight/survivability screening for a species with a reasonably short vegetative period. Reproductive success relates closely to the available nutrient supplies within the plant and thus should be subject to alterations caused by these losses. However, additional research is needed to establish a reproducible protocol for reproductive success for more than one species. A reproductive success test might be more appropriate for a Level 2 or 3 assessment tool. It was argued that some endpoints may be too variable in the response, e.g. photosynthesis.

How can individual plant parameters be used to predict population effects?

The Panel noted that Levels 1 and 2 are limited greenhouse toxicity screening tests. The extrapolation from individual to population endpoints is fraught with problems and uncertainties, particularly for endpoints that lack a basis in functional population biology. Any transfer from individual to population endpoints requires a modeling approach that incorporates species-specific attributes such as annual versus perennial, dispersal mechanisms, safe sites for seed germination, etc. Such information is well beyond the scope of a screening exercise and might better fit into Level 3 or Level 4.

Question 5 **<u>Consideration of physiological and biochemical markers</u>**
In addition to established measurement endpoints (gross acute), recent literature suggests it is important to consider alternative endpoints pertaining to physiological and biochemical effects (e.g., O_2 production, carbon fixation, etc.). The endpoints assessed in seedling emergence and vegetative vigor tests may not fully detect physiological changes that are detrimental to plants.

How important is it to consider physiological or biochemical endpoints when determining chemical risk to plants?

The Panel cannot support the use of physiological or biochemical endpoints in routine assessments until additional research is conducted to demonstrate their relationship to survival, growth, yield, or reproduction. There are three types of biomarkers: those that indicate exposure, those that measure effects (that should be related directly to fitness), and those that elucidate

mode of action (and allow extrapolation to untested species). Only biomarkers directly related to fitness should be considered in Level 1 . However, it is unclear at this time if such correlations currently are available to allow selection of appropriate measures. Most members of the Panel agreed that although there is a large number of methods to assess carbon fixation and enzyme activity, variation at the leaf level or even the whole plant level is so great that measurements of carbon fixation or enzyme activity are not suitable for use in Level 1 at this time. To achieve use at Level 1, one would require a large number of samples per assay or measurements of whole sections of canopy under constant conditions. Questions about relevancy, variability, and cost of such assays require further discussion. What would be the advantages and disadvantages of replacing seedling emergence and vegetative vigor endpoints with physiological and biochemical endpoints? Before adding any of these endpoints to the standard test methods, the Agency should convene a Work Group of biomarker experts to discuss the state-of-the art, review various protocols, and consider the applicability of such endpoints to ecological and agronomic endpoints.

It is the Panel's recommendation that biomarker endpoints be selected that directly underpin growth and development. The following three classes of biomarkers may be best suited for this objective:

a. Biomarkers based on whole-plant physiology (versus organelle physiology);
b. Biomarkers detectable using HSI, a remote sensing technology (hyperspectral); and
c. Biomarkers based on genomics.

Physiological tests appropriate to whole-plant physiology and stress ecology include:

a. Net photosynthesis;
b. Stomatal conductance to H_2O vapor;
c. Ethylene emission; and
d. Biomass partitioning by organ (e.g., dry weight partitioned to roots, stems and leaves).

As reference material for the above physiological properties, the following are noted:

Lambers, H., FS Chapin III, and TL Pons. 1998. *Plant Physiological Ecology*, Springer, New York.

Nobel, PS. 1999. *Physicochemical and Environmental Plant Physiology.* Academic Press, New York

Key journals for up-to-date methods in plant physiological ecology include *Plant, Cell and Environment*; *New Phytologist*; *Physiologia Plantarum*; *Plant Physiology*; *Canadian Journal of Plant Science;* and *Tree Physiology*.

What physiological or biochemical changes would be considered adverse? Please quantify (e.g., 20%, 50%, etc.).

The EC25 appears to be an appropriate trigger assuming there are a limited number of endpoints and the endpoints are functionally selected. It is noteworthy that the trigger is highly susceptible

to the exposure methodology, requiring attention to exposure-response issues.

What would be the advantages and disadvantages of replacing seedling emergence and vegetative vigor endpoints with physiological and biochemical endpoints?

There are very limited reasons for continuing to conduct testing using solely seed germination, seedling elongation and growth. In fact, there are several valid reasons why these protocols may be misleading. The Panel recommends that more functionally based methodologies be developed as complementary tools. The most immediate need is for studies based on whole-plant physiology.

Which additional endpoints could be used to assess hazard to non-target plants?

The Panel suggests the following endpoints to assess hazards to non-target plants:

Whole plant physiology, including net photosynthesis, stomatal conductance, and photosynthate allocation (e.g., loading, etc.)

Remote sensing using HSI technology that can be hand-held, airborne or satellite positioned. If these studies are developed properly, the HSI technology has the potential to be highly diagnostic, capable of pinpointing chemicals in foliage, and resulting effects on a plant's physiological vigor.

Genomics, including emerging methods to evaluate the potential for new chemicals to regulate functional segments of the genome that underpin critical plant processes (e.g., carbon fixation, respiration, phloem loading, etc.).

Community Importance index (CI) proposed to be used for detecting keystone species could be used (Power et al., 1996).

In summary, variability among plants is great and the emerging capabilities of biomarkers and whole plant physiology hold great promise. There are numerous places where new physiological information could be applied to current data.

Question 6 <u>**Aquatic and terrestrial reproductive effects**</u>

Because existing plant toxicity tests do not provide an adequate trigger for reproductive testing, the PMRA and the EPA propose that reproductive testing be conducted at Level 1. For aquatic plant assessments, we propose the testing of two species (rice and nodding smartweed). For terrestrial plant assessments, we propose two of three test species [cherry, mouse-ear cress (*Arabidopsis thaliana*), and canola (*Brassica rapa*)]. The agencies request guidance on how to refine the assessment of reproductively sensitive plant species.

What are the advantages and disadvantages of the proposal to assess reproductive effects at Level 1 versus higher levels?

The Panel does not support the incorporation of reproductive effects at Level 1. Level 1 should continue to focus on easily obtained and interpretable endpoints such as seedling growth and vegetative vigor. However, one advantage to assessing reproductive effects is that the chance of overlooking an important effect at Level 1 for rooted aquatic and terrestrial plants is lowered. This could be an important advantage that outweighs the disadvantage of the time and effort required to generate the data, for cases in which reproductive effects have been documented in the field. It is noted that this is a controversial conclusion, based on only a few species and herbicides. The proposal document indicates that these effects were seen at levels "far less" than the recommended application rates. However, it is important to remember that reproduction in plants is highly conserved, and that early seedling growth is generally more sensitive to effects of chemicals than are reproductive effects.

The reported incidence of crop damage may support the idea of not including reproduction tests at Level 1. Given the extensive use of pesticides in North America, the fact that only 1,010 adverse incidences were reported may be a sign that the process is relatively effective. Most incidences of non-target plant damage can be attributed to inappropriate practices of pesticide application. Inadvertent exposure to pesticide spray drift generally is associated with applying pesticides using inappropriate equipment (e.g., incorrect spray applicator, nozzle, pressure, etc.), when meteorological conditions that are not conducive to application (e.g., wind conditions are inappropriate), or failure to respect buffer zones or label information.

Reproduction tests are relatively long-term in nature. The longer the test duration, the greater the probability of confounding factors or problems arising. These factors include a) interactions between the test compound and soil matrix, b) increased likelihood of pathogens causing damage or disease, c) nutrient limitations in substrate, d) accumulation of salts in substrate, e) influence of degradative mechanisms on pest control product; etc.

Rather than require reproductive testing at Level 1, there should be an option available if a flag is raised during problem formulation. In particular, if the mode of action suggests that reproductive effects could occur, this test should be incorporated into Level 1. Two test protocols are currently available but further validation is needed. The Panel strongly supports research to better understand if reproductive effects occur in the absence of other, currently measured effects.

What are the advantages and disadvantages of partial- versus full-life cycle testing?

A partial life-cycle test might be more practical, less costly, and require less time to complete (depending on the species). However, the critical window for effects might be relatively short and not adequately represented by a partial life-cycle test. A full life-cycle test is more likely to elucidate adverse reproduction effects, and hence, be more reliable. Disadvantages of the full life-cycle test include the lack of data on performance criteria and no F1 seed viability endpoint.

As a result uncertainty remains regarding the effects on populations. Although more costly and longer, full life-cycle tests allow for optimal production of age- or life cycle stage-specific survival and reproductive estimates (i.e., allow demographic projections of population consequences).

Reproductive endpoints could be important because they directly relate to population growth rate and fitness. There are problems associated with measuring reproductive endpoints, not the least of which is time and subsequent expense to get the measure. However, physiological measures that can be shown to correlate with fitness or reproductive potential of plants should be included at Level 1 to provide the sensitivity needed to detect subtle effects of herbicides not detected with conventional emergence and seedling vigor measures. Thus, we have defined a research objective of correlating more efficient physiological methods with sexual and asexual reproduction measures.

For full life-cycle testing, what are the advantages and disadvantages of testing mouse-ear cress versus canola?

From a practical perspective, there are advantages to using mouse-ear cress over canola. The test is shorter (36-46 d), the species genome has been described and is documented, and the variability can be reduced by using clones. As a research tool, cress has the advantage of use of selective transgenics to elucidate the effects of products with specific modes of action. The canola tests may have more background information available to draw upon during data interpretation. The test with canola is longer (60-120 d) and the source of seed is limited. However, as a member of the mustard family, it also may represent cereal or grain species.

What are some alternative approaches and species that address potential reproductive effects?

It would be very helpful to generate age- or life-cycle stage-specific survival and reproductive information amenable to demographic analysis. Use of such vital rates in demographic models is the approach taken in biological sciences to predict characteristics and consequences to populations. Individual-focused information allows only gross and potentially incorrect predictions of population fate.

What reproductive testing endpoints (e.g., pollen viability, seed formation, etc.) are critical for Level progression and for assessing chemical risk to non-target plants?

Until the problems discussed above are addressed, it is inappropriate to speculate which reproductive testing endpoints would be critical for either Level progression or for assessing chemical risk to non-target plants.

What reproductive EC values (e.g., NOAEC, EC_{05}) are critical for Level progression and for assessing the risk to non-target plants?

The endpoint used for Level progression and to assess the risk to non-target plants depends on the assessment endpoint (i.e., what it is one is trying to protect). This must be decided during the problem formulation and the iterative process that results in the development of the conceptual model. In the absence of data needed for conventional population-level predictions, it is difficult to assess which metric is best. An EC5 probably is too difficult to measure. The present use of hypothesis tests to generate NOAEC data has been criticized due to fundamental statistical flaws. If restricted to these options, the NOAEC (or roughly, an EC25) could be used.

Other methods are applicable, such as the regression method of Liber et al.,1992, the least significant difference statistics from conventional hypothesis tests, and bioequivalence testing.

How critical is it to consider both modes of reproduction (sexual versus asexual)?

Different species rely to different degrees on sexual and asexual reproduction to assure population persistence. The decision about which process should be considered most closely must be driven by the problem identified for the particular chemical/formulation and exposure scenario(s).

Question 7 **<u>Marine algal toxicity testing</u>**

The range of response among different marine algae can be as great as is observed in freshwater species. Currently, only one marine algal species test is required at Level 1. Researchers have recommended a test battery which includes species of diatoms, green algae, and dinoflagellates as well as golden-brown algae. This test battery should provide a range of responses until sufficient comparative toxicity studies are available to determine sensitivity ratios. In addition to the currently required marine diatom, the proposed plant testing scheme requires the testing of one marine algal species in each of three previously unrepresented Divisions (Chyrsophycophyta, Pyrrhophycophyta, and Rhodophyta).

Are the four proposed marine algal species representative? What other marine algal species should be considered? Why?

The question as stated needs clarification with respect to the term “representative.” As noted in the Agency’s proposal document, many tons of various herbicides end up in South Florida estuaries, and herbicides are present in the vast majority of sediment samples from the Atlantic and Gulf Coast areas. Algal species that predominate in areas of high potential exposure should rank high on the “should-be-considered” list. The Panel reiterates that problem formulation resulting in a conceptual model should direct the assessment. Marine species are not necessary if there is little chance of marine exposure. The Agency should conduct further assessments to determine the relative sensitivity of marine versus freshwater species. The fact that the range of

sensitivity of marine species is as great as that of freshwater species is not necessarily a reason to test them if the range of current freshwater species will bracket that of marine species. It should be mentioned that additional information concerning concentrations of herbicides in estuaries is needed.

Skeletonema costatum	This is a popular diatom (527 hits, using the Science Citation search engine), and seems to be more nutritious than *Phaeodactylum*. The justification for using *Phaeodactylum*, if *Skeletonema* will be used, is not clear. If algal test species are being selected with consideration to ecological function, it might be advantageous to consider species such as *Dunaliella tertiolecta* (388 hits by Science Citation search). *Skeletonema* is easy to culture; *Thalassiosira pseudonana* is also a reasonable diatom to select for testing.
Gonyaulax polyedra or Pyrocyctis lunula	One problem with the bioluminescent marine dinoflagellates tests is that inhibition of light production is not a guarantee for effects to any other critical function (carbon fixation, motility, protein synthesis, etc.). It is doubtful that the Agency is interested in protecting marine systems on the basis of light production, unless obligatory, strong correlations are known between light production and more fundamental parameters of interest. Good alternative parameters that can be measured exist: current technology allows fairly easy determination of cell number and cell size, for algae in culture. These two properties, measured through time, are more fundamental than light production. Other relevant parameters could include the relative proportions of carbohydrates, lipids, and proteins or the relative proportions of photosynthetic and accessory pigments. The latter can be quantified rapidly at low cost by High Performance Liquid Chromatography.
Champia parvula	The Red algal reproduction test (successful gamete formation) is more easily used to predict consequences.
Phaeodactylum tricornutum	The Golden brown algal species is convenient and reasonable. However, it is not a “choice” food item for copepods (see Lacoste et al. 2001; for example) so justification for the use of *Phaeodactylum* should be sought in a non-food-web context. Many studies have been published for this diatom (563 hits, using the Science Citation search engine), which makes it attractive as a test organism.

Could a macrophyte growth (Ulva?) assay be useful? The Panel suggests consulting with the

Agency's marine specialists for additional input on important species.

What are the limitations of the protocols and the availability of the proposed marine algal species?

Most of these species can be obtained and cultured readily. However, protocols are not presented with performance criteria, and there is no mention of reference toxicants. These are generic qualities of a standard test protocol that are needed. Tests might be reserved for Level II once a conceptual model that includes marine species exposures is developed.

The endpoints are logically distinct and their relative ease of use to determine the presence or absence of "unreasonable risk" is quite different. How does one predict adverse consequences to individual fitness or population persistence from some of these metrics?

It would be a logical challenge to use these metrics in a probabilistic approach in which all numbers are treated equal. For example, algal bioluminescence inhibition and gamete formation are not equal in their consequence on species persistence under the influence of a stressor.

Question 8 **<u>Aerial exposure testing of floating aquatic plants</u>**
Chemical exposure can occur from drift deposited onto the leaf surfaces of floating plants. For contact toxicants, it is not sufficient to test only aquatic exposure (i.e., chemical dissolved in water). It has been observed that the sensitivity of *Lemna* to a contact herbicide can increase several-fold with a foliar exposure compared to the conventional exposure through the growth medium. Although there are limited data on aerial exposure, both Agencies believe that this type of study should be conducted on a routine basis.

What are the pros and cons of routinely requiring foliar exposure tests for floating aquatic plants?

The clear benefit of foliar exposure tests on floating aquatic plants is that another potential exposure option would be studied closely. However, depending on whether the toxicity testing context shifts to generating a clear problem formulation step, this routine requirement may result in time and effort being spent on testing a less appropriate exposure scenario. The tests would increase the confidence level. Since aquatics have different mechanisms to enable them to compete in an aquatic environment, there might be different effects on them. Additional tests could be deemed excessive due to fast competitive regrowth that typically occurs in water.

How can the methodology by Lockhart *et al.* (1989) be modified for future testing requirements?

This methodology utilizes overspray so that the foliar parts of the plants and the water were treated at the same time. This method has the advantage of being able to be computerized. It is

recommended that the Lockhart metholodogy be used together with exposure by the water column only.

What research needs can be identified for foliar exposure testing?

Since only one species of aquatic macrophyte is currently tested, it is reasonable to test several different genera to select the most appropriate species and develop a standard methodology. Testing on Lemna alone does not provide a balanced representation of the aquatic plant flora commonly found in the water bodies of North America. Evaluations should develop information on the ability of species to represent the range of plant types in the aquatic environment. Protocols, when possible, should emphasize ease of culturing as an important aspect of testing. It should be noted that ferns are ecologically important in North America.

Question 9 **<u>Submersed aquatic vascular plant testing</u>**
Submersed aquatic vascular plants are morphologically different compared to terrestrial vascular plants or algae. Currently, *Lemna* spp. (floating) are used to predict effects on submersed aquatic vascular plants. As *Lemn*a is a monocotyledon, it would be preferable to also require a representative dicotyledon. The agencies recommend *Myriophyllum* spp. to represent dicotyledonous submersed aquatic species.

What is the SAP's opinion on the proposed toxicity test requirement with a submersed aquatic vascular species?

It was generally the view that all growth habits and groups of plants should be considered. Additional test species should be added judicially based on existing data. One Panel member stated that the addition of a submerged macrophyte under the Level 1 testing regime will increase the breadth, ecological relevance, and ultimate quality of herbicide risk assessments.

What are the Panel's thoughts on the selection of *Myriophyllum* spp. to represent dicotyledonous submersed aquatic vascular plants?

An aquatic dicot is needed in the test array for detection and evaluation of herbicides that have auxin-simulating properties. Myriophyllum is a logical choice since there has been a significant amount of work conducted with this species. It is widely distributed; easy to culture; and is a frequent target of aquatic weed management programs. Thus, there is a large amount of data that exists concerning field-level responses to a wide variety of herbicides. This provides a unique opportunity to develop more extensive datasets on a model species that can be used to establish the full spectrum of plant responses to herbicides ranging from short-term physiological responses to phytostatic to phytocidal responses. However, there is evidence that Myriophyllum is not as sensitive to some herbicides (e.g., acetanilides) as other plants (Fairchild et al. 1988) and there is little information concerning its relative sensitivity compared to other dicotyledonous aquatic plants.

The Panel also noted that although *Myriophyllum* seems to be an appropriate choice, other genera should not be neglected.

What other monocotyledonous and dicotyledonous submersed species should be considered? Why?

Other monocot species that should be considered include *Elodea* sp. and *Egeria* sp. Both of these can be grown from cuttings which makes them amenable to culture and testing. *Najas* sp. can be easily germinated from stored sediments. In this case they establish roots that could be used to determine the availability of contaminants from sediments.

Other dicot species for consideration include *Ceratophyllum* sp. All of the above species are ecologically relevant due to their wide distribution; are available commercially; and are easily cultured in the laboratory.

Of these species *Najas* sp. (monocot) and *Ceratophyllum* sp. (dicot) have been shown to be especially sensitive to herbicides. In a study that compared the sensitivity of 11 aquatic plants (6 species of algae and 5 species of macrophytes) to four herbicides (atrazine, metribuzin, alachlor, and metolachlor) they were the two most sensitive macrophytes and were consistently in the top 25% of all species in relative sensitivity (Fairchild et al. 1998).

There are many other rooted species of ecological importance that could be tested including *Vallisneria* and *Potemogeton*. However, species within these genera frequently have a large rhizome that serves as an energy reserve and decreases the sensitivity of the plant to some herbicides. Furthermore, the presence of the rhizome can increase the dispersion of weight data in a dataset. Therefore, they should be considered with appropriate caution.

A marine species such as *Ruppia* or the sea-grass *Thallassia* should be considered. Due to recent concerns about pollution and sea-grass health it would be very important to select one of the sea-grasses for review.

Basically, one species in each different group should be examined. Just as was true concerning species sensitivity among terrestrial plants, much research is needed to better understand this phenomenon for aquatic species.

Question 10 **<u>Emergent aquatic vascular plant testing</u>**

Currently, *Lemna* spp. (floating) and terrestrial vascular plants are used to predict the effects of toxicants on emergent aquatic vascular plants. However, there are no available data to support the use of these as surrogates for emergent aquatic plants. Many emergent aquatic plant species prosper in both aquatic and terrestrial habitats. In an aquatic exposure scenario, emergents are unique in that they can be exposed *via* root or stem uptake of contaminated water in addition to foliar exposure from an over-spray.

What are the Panel's thoughts regarding the physiological differences between emergent aquatic and terrestrial vascular species or *Lemna* sp. to support the proposed approach?

It is debatable how well terrestrial species might represent emergent aquatic species, from a physiological perspective. There clearly is a gradient of biochemical and physiological properties, ranging from "purely aquatic" to "practically terrestrial", even considering only those species classified as emergent. Emergent species that support substantial below-ground biomass structures often depend critically on gas-flow through leaves and into roots, and this ability is not well expressed by terrestrial species. The tolerance of some aquatic emergent species to anaerobic conditions is allowed by metabolic pathways that differ in balance from those used by terrestrial plants with less below-ground biomass to support and fewer reducing conditions to contend with.

For these reasons, *Lemna* would not be a good model for floating-leaved (attached) aquatic plant species such as *Potamogeton natans*, and terrestrial plant species probably would not adequately represent responses of emergent aquatic plant species. Also, because *Lemna* is a floating plant species with limited root structure it is probably not a good species to represent aquatic emergent plants, although no comparative data exist upon which to base a conclusion.

What are the Panel's thoughts on the selection of rice and nodding smartweed to represent emergent aquatic vascular plants?

Both rice and nodding smartweed are reasonable choices. There is a contradiction with using only two species to represent emergent aquatic vascular plants, considering that 10 species seem insufficient to represent terrestrial vascular species, and that aquatic habitats can function as "collection points" for chemicals of agricultural significance.

What other emergent aquatic vascular plant species can the Panel recommend to satisfy both terrestrial and aquatic testing requirements?

First, data on rice normally not submitted to the agency for review should be made available. Second, Bidens cernua, Mimulus ringens, Echinochloa crus-galli, as well as some fern species, have been used successfully in greenhouse experiments. Pickerel Weed (*Pontederia cordata*) and Water Grass (*Hydrochloa caroliniensis*, now *Luziola fluitans*) are monocots that will survive when water is withdrawn as long as the soil is moist. Water Grass actually assumes a different growth form when out of water. Water-lily (*Nymphaea*) and Spatter-dock (*Nuphar*) are both good dicot candidates with their large floating leaves. It should be made clear that recommendations for other species are useful but the methods for testing may require species-specific modifications, and performance criteria for the species must be developed before the species can be used in a regulatory framework.

How important is it to consider routes of exposure other than foliar exposure to emergent aquatic species (i.e., absorption from the water column by the submersed stem or from sediments *via* the roots)?

Since at least some herbicides can accumulate in sediments, the possibility of damage by exposure from sediment-borne chemical probably cannot be excluded. Foliar exposure is probably the most likely route of damage, followed by sediment exposure, followed by water exposure. Perhaps the smartweed test could be modified to include a foliar-exposure component, and a sediment-exposure component, conducted separately, to cover these bases. It is noted that not all pesticides enter water systems from spray drift or direct application. Surface water runoff of agriculture lands to which pesticides have been applied is an important consideration, as well.

What is the Panel's opinion regarding extrapolation from terrestrial vascular species or *Lemna* spp. to emergent rooted aquatic vascular plants?

See above.

Question 11 **<u>Mode of action</u>**

The agencies are proposing a seedling emergence and a vegetative vigor test at Level 1. Various factors contribute to eliciting toxicity of a particular chemical (e.g., mechanism of uptake by a plant, the mode of action of the chemical within a plant, application parameters, etc.). Information on these factors could reduce the need to conduct a seedling emergence or vegetative vigor test.

This initiative has a great deal of promise. The data on mode of action come from the manufacturer and probably focus on one primary mode of action. There is reason to believe that many chemicals have multiple modes of action. Therefore, even if a primary mode of action is reported, it is prudent to recognize the potential for other modes of action. Accordingly, Level 1 testing should be conducted independent of the mode of action. There are opportunities at higher levels for restricting the tests in light of the data collected at Level 1, and this is the greatest opportunity for the application of "waivers."

What criteria might we use for data waivers based on the mode of uptake and the mode of action of a chemical?

There are a number of issues that might be considered as candidates for data waivers. For example, Level 1 or 2 tests might include approaches based on the vector for dispersal. Chemicals whose composition precludes foliar uptake should not be evaluated for that mode of exposure. Chemicals that have a half-life measured in days are not likely candidates for bioaccumulation or fate studies in an ecosystem context. If the chemical is sprayed on landscapes and has a very short half-life, root tests are unlikely to be valuable in evaluating toxicity.

However, the best place for adoption of "waivers" is at levels above Level 1. If information is needed on the range of species sensitivity, it might be advisable to test even the species known to be relatively resistant to the chemical. The mode of action becomes more critical in higher Levels

(especially Level 3) when focusing the risk assessment on specific species and endpoints. Understanding of the mode of action provides a higher level of knowledge and information to focus and underpin risk or managment discussions.

How could the number and types of tests be reduced with respect to application parameters (i.e., timing, method, etc.)?

See the above discussion for some preliminary issues.

One major evaluative tool that may short circuit further tests is the application of remote sensing. In preliminary field tests under realistic conditions, if remote sensing is incapable of detecting effects or the chemical toxins, further tests that are ecologically based (e.g., bioaccumulation or multiple species) may be unnecessary. This has application at Levels 2 and 3.

What other factors might reduce the amount of testing required?

There is a need to better characterize the degree to which microbes chemically degrade pesticides in the field. If tests in the laboratory demonstrate that an active soil microbial mass can shorten the half-life of a chemical, fate and bioaccumulation studies in the field may not be as critical.

Question 12 **Terrestrial Species**

Both Agencies have recognized the need to consider sensitivity to chemicals among a broad range of ecologically-relevant plant families. The Agencies have proposed to increase the number of families tested to reduce uncertainty and variability with respect to sensitivity. Researchers have recommend a test battery including non-crop and woody species to encompass a range of response until sufficient comparative toxicity studies are available to determine sensitivity ratios. The selection of the families and the species within those families was based on the feasibility of using them as test species and their economic or ecological importance.

What are the Panel's thoughts on the proposed terrestrial species at Level 1?

The Panel agreed that Level 1 testing is a screening level assessment. As such, increasing the number of species representing a larger number of families may give breadth to the data set. However, there are no data demonstrating that this data set would be more sensitive, and since the result of Level 1 is a risk quotient, the most sensitive species drives the toxicity number. Species selected must have performance criteria. Research must be targeted to develop the appropriate suite of plant test species. This should include a characterization of the uncertainty that exists with the present set of plant tests.

The test species battery proposed for Level 1 is more comprehensive than the current test battery; however, the rationale or scientific justification for selection of these species is not obvious nor is

the need for additional species convincing. Risk assessment is predicated on probability and the process benefits from robust species sensitivity distributions as well as the distributions of concentrations likely to occur in the environment. More test species and more tests to generate more data will not necessarily translate to better decisions regarding risk. If the species test battery that currently exists adequately represents the range of sensitivities of the 26 species proposed, then there is likely no need to expand the test battery beyond the 10 species. However, there is a major advantage of having a choice among a larger number of potential test species, and the selection of a subset of the 26 species could be part of the problem formulation.

Important considerations for species selection include how, where, when, and for what purpose the pest control product would be used as well as how the pest control product would be applied (pre- or post-emergent, soil incorporation, formulation). The physical-chemical characteristics of the product (e.g., solubility, adsorption coefficient, volatility, persistence, susceptibility to microbial degradation, photolytic breakdown or transformation, etc.), the matrix to which it is being applied, and the identification of potential and realized exposure pathways are also important.

Non-crop and woody species have been added to the list of test species with little rationale provided other than that they represent groups of organisms not previously represented taxonomically. If this is the only criterion for expanding the list, then epiphytes, liverworts, mosses, lichens, CAM metabolizing plants, desert species, tundra species, and orchids (among others) should also be considered.

The following process, used to select a species test battery for two regulatory initiatives, was proposed by one Panel member:

> One regulatory initiative was the development of soil quality criteria; the objective was to select a test battery of very sensitive species. The other was a generic test species battery for ecological risk assessment; the objective was to select a plant species with sensitivities to soil contaminants that were representative of the range of sensitivities that might exist in a "natural" plant community. The process used to select these two test batteries is briefly outlined below. It is a different approach than the EPA/PMRA has proposed but it is not necessarily better.
>
> A number of *a priori* selection criteria were identified and used to select the test battery. Some proved more useful than others when the selection process was actually implemented. In addition to the *a priori* criteria, there were several *a postori* selection criteria. At the end of this process, three questions were answered; these questions served as a reality check for an academic process. The details for this exercise can be found in an Environment Canada report that is available from Rick Scroggins, Test Method and Applications Section, Environment Canada, Environmental Technology Centre, 3439 River Road, Ottawa, Ontario K1A 0H3 (613-990-8569).
>
> The weighted *a priori* selection criteria were a combination of functional and taxonomic characteristics. They are listed below along with the assigned weighting factors.

Criteria	Weighting Factor
Time to germination/emergence	2
Type of germination (epigeal/hypogeal)	1
Seed pretreatment	5*
Crop vs non-crop species	2
Monocotyledonous vs dicotyledonous species	1
Nature of photosynthetic system (C3 vs C4)	2
Source, availability, and quality of seed	5*
Relative sensitivities to known contaminants	5
Critical variable requirements (pH, nutrients, mycorrhizae)	5*
Above- and below-ground crop species	3
Type and nature of root formation	5*
Phenology and life-history characteristics	1

* These criteria were not used in the final assessment

> One of the problems encountered was that of information on the sensitivity of different plant species to a common contaminant. These data were not readily available, so seedling emergence test methods were used to determine the relative sensitivities of 30 potential test species to a reference toxicant (boric acid) in two soil types. The data demonstrated that there could be a difference of greater than three orders of magnitude in the sensitivities of these species to a single toxicant. Using a number of toxicity criteria (e.g., % emergence, LC50s, LC20s, LOAECs and NOAECs for shoot and root lengths, LOAECs and NOAECs for shoot and root wet and dry masses) along with some practical considerations (e.g., time to emergence, test duration required for sufficient biomass for reliable measurement, ease of root separation from soil, time to seedling emergence), species were ranked according to their relative sensitivities to boric acid in soils. A decision analysis was used to integrate the selection criteria with the toxicity criteria to derive species that would be "representative." When species appeared equal, three important practical questions were asked: 1) did the potential test species exhibit >70% emergence in the control soils?; 2) did the potential test species exhibit a concentration-response relationship with the contaminant in soil?; and 3) was sufficient biomass present at the end of the test to produce reliable metrics? The answers to the questions enabled the separation of species with comparable relative rankings and resulted in recommendations regarding the suitability of the candidate test species.

"Sensitivity" of plants to contaminants is really a moot point. There are few generalities that can be made regarding this issue. Patterns of sensitivity are rare and sensitivity of plants to contaminants can be soil specific, species specific, contaminant specific, endpoint specific, and even test duration dependent. Changing any of these variables can influence the interpretation of comparative sensitivity. EPA/PMRA would have to examine the range of sensitivities of the proposed 26 test species to determine if it differs substantially when compared to that of the current 10 test species. The comparison should include an analysis of the variability and uncertainty associated with the evaluation. In the final analysis, there must be a sound scientific

rationale for adding more species to the initial test screen.

Below are more detailed comments from Panel members about the proposed list and potential new terrestrial plants for screening assessment:

MONOCOTS:
Grass Family – corn + rye grass + one other native species
The grass family is without doubt the most important plant family in the world. With its world wide distribution, size, and variety of growth forms, attention should be given to a review of more species . . . certainly at least one species in each subfamily.

Sedge Family – purple nut sedge
Comments concern the great importance of Sedges in the habitats in the northern latitudes and that there are many species that lack rhizomes and tubers. Yellow Nut Sedge was mentioned as more important in North America than Purple Nut Sedge. Both Yellow Nut Sedge and Purple Nut Sedge have rhizomes and tubers. Purple Nut Sedge is difficult to propagate for research work while the tubers of Yellow Nut Sedge, an edible food source for humans and wildlife, will last for long periods without much care.

Lily Family – onion
There are other important non-bulbous species.

The Orchid Family (Orchidaceae)
This is arguably the largest plant family in the world. Many species of Orchids are epiphytic as are most species in the Bromeliaceae (Air Plant Family).

Pteridophytes (Ferns)
It would be desirable to add a Fern or two to the list of species requiring additional research and/or protocols.

Orchids, Air Plants and Ferns have large advocacy groups. Sensitivity of ferns can be an issue; testing with orchids and epiphytes would require method development and would meet with a great deal of resistance for specific organizations devoted to the propagation of these plants.

DICOTS:
Aster Family – lettuce + one other
A huge family of great importance to mankind. It would be important to review a species from each section of the family.

Beech Family – oak
Certainly in North America, Oak is the genus of choice. Carya illinoinensis (Pecan) in a closely related family, the Juglandaceae (Walnut Family), could also be a wise choice.

Pine Family – sugar or loblolly pine

Pine is the obvious gymnosperm to research; however, Cypress in the Cupressaceae (Cedar Family) also could be a good common species and family.

Mustard Family – canola + Arabidopsis
Another important family for food and weeds.

Pea or Bean Family – soybean + yellow sweet clover + one other
Another very important and very large family that should require more emphasis by research on several other species in other sections of the family.

Morning-glory Family – 1 species
There are different modes of action among species in the same genus (Ipomoea). It would be desirable to investigate this.

Other groups of plants that are notable and lacking are the Horsetails (a Pteridophyte) and Mosses and Liverworts.

What (if any) additional species or groups are not adequately represented in the proposed testing scheme at Level 1?

There currently is a need for woody perennials and tropical plant species on the test list. This may be of use in light of population shift of the United States to the southern states and the potential for alterations in crop selection within these areas based on perceived "global change" scenarios.

Are there better approaches for selection of species besides the taxonomic / phylogenetic approach (e.g., ecological or functional approach)? What are they?

One approach would be that of physiological sensitivity to various classes of pesticides. They are not restricted to taxonomic relations and efforts should be initiated to identify suitable types. This approach may also be enhanced through the inclusion of ecologically important native plants for those regions in which the product will be primarily used.

How can the agencies improve their knowledge of the variability in sensitivity of the proposed test species?

One answer to this question is the initiation of a research effort to identify and develop standardized tests for species of higher sensitivity to ensure a conservative result from the Level 1tests.

Question 13 <u>Additional Species Testing</u>

Level 2 is envisioned as primarily an assessment level that utilizes refined exposure methods and toxicity assessment. However, additional testing may be needed to clarify uncertainties before advancing to Level 3, such as laboratory to field extrapolation or specific dose-response curves. Level 3 is envisioned to include expansion of testing in two areas: reproductive testing and acute testing of keystone species. Keystone species can be selected in a couple of ways: (1) keystone species within families triggered by risk identified at Level 2; and (2) keystone species within new families that are within a structure/function group (e.g., woody plants) identified to be at risk at Level 2 or identified in incident reports.

What are the Panel's thoughts on additional testing to clarify uncertainties on previously tested species in Level 2?

Selection of additional test species is dependent on a thorough problem formulation to trigger the transition into ecological risk assessment. The movement into higher levels of testing should be iterative in nature, increasingly focusing on more sensitive and important species. There is a concern that already many species are to be tested, so any additional species needs to be carefully assessed and guided by any previous knowledge (e.g., mode of action, proposed regions for release of the chemical, etc.).

Under the assumption that Level 2 testing is a transition into the realm of risk assessment and has the background information of a generic sensitivity test that has triggered the need for Level 2 testing, efforts can focus on better characterization of exposure scenarios. Thus, once the potential regional use patterns are established and specific areas vulnerable to exposure are identified for use of the pesticide or chemical, a list of test species for those regions or areas could be identified. The list of species should include crops of major economic importance, threatened and endangered species, dominant species, keystone species, and important functional groups. Functional group diversity, rather than taxonomic group diversity, should be emphasized in species selection. These additional test species would be tested in Level 3.

The test species response measures should be sensitive enough to determine if significant ecological changes occur to the plant, population, community or ecosystem following exposure to the chemical at expected concentrations. There may be very subtle stresses on a functionally important species imposed by the chemical that does not cause obvious symptoms (acute or even reproductive effects) but which might cause a competitive shift that allows compositional changes in the community. Thus an important point is to select critical species for additional testing and use measurements that will elucidate even subtle impacts or allow triggers to move to Level 4 where macrocosm studies may be the preferred approach. In addition, experiments and measures must be selected to deliver probabilistic outcomes for risk assessment. Consequently, if a test species at Level 1 shows a phytotoxic response, other species should be tested. Other species may be chosen based on the following considerations:

- <u>An adequate representative sensitivity distribution not established at Level 1.</u> This allows for Level 1 to use the minimum number of species. If a full sensitivity range is not obtained, the

extra species needed to extend the range can be added.

- Regional specificity in species or families not covered by Level 1 test species. This allows for adding species of ecological importance to a specific region where the chemical will be used. For example, if a herbicide is targeted for use in the Northern Great Plains, Great Basin, or Rocky Mountain regions, a test species from the family Salicaceae should be included among additional test species, because they dominate most all riparian areas which are critical sources for species diversity.
- Threatened and Endangered plant species that occur in the region of exposure should be incuded to determine their sensitivity at Level 3.
- Dominant, keystone and important functional group species. Representatives of these species can be included in Level 3 testing to increase certainty of tests.

What are the Panel's thoughts on having two areas of focus in Level 3 (reproductive testing and acute testing of keystone species)?

Reproductive testing is more of a direct measure of fitness, but may want to include offspring viability as well. Acute (visual injury) measures maybe used in a stepwise approach where injury does not allow a species to get to reproductive stage. There are valid reasons for investigating reproductive effects although the trigger used to shift to this effort in Level 3 is unclear. Only a small proportion of the chemicals are likely to have effects on reproduction, so a strong biological basis for suspecting such effects needs to be developed. Assuming the triggering is accurate, this effort is warranted.

Additional species testing (Level 3) should be directed toward understanding the distribution of sensitivities within plant groups that have been shown to be sensitive to the herbicide of concern. This may be a taxonomic group, a physiologic group, or a functional group. It should be focused on species that are present in the area of intended use and must be designed to answer very specific questions. An *a priori* description of how the additional information will be used in the assessment process should be provided as reassurance that the request is meaningful and will be useful in reducing uncertainty. If there are particular species of concern in the area of intended use (e.g., dominant species, threatened or endangered species, or species of particular use by a desired wildlife species) the degree of certainty about the potential hazard to these species should be high. The information available should be reviewed with this in mind, to determine if the species sensitivity distribution is robust enough to capture these concerns, if a similar species has been tested, or what is known about mechanisms of action will suffice. In regard to additional endpoints (e.g., reproductive endpoints), these should be tested only if it is known that 1) they are more sensitive than the current measurement endpoints or 2) the mechanism of action is such that the herbicide directly targets a reproductive endpoint. However, it has been shown that seedling vigor often is a more sensitive endpoint that either sexual or asexual reproduction. Furthermore, it has been suggested that many of the physiological measures (e.g., photosynthesis) may have equal or greater sensitivity than reproduction. In these cases, there would be no need to test reproductive endpoints, because risk assessment and management decisions will be sufficiently conservative to protect this portion of the plant life cycle. Certainly, the underpinning science

needs to be researched.

What are the advantages and disadvantages of additional species testing at Level 3? Should the additional species be focused on keystone or ecologically significant species prevalent in areas of chemical use?

Other species, regardless of their known function in the ecosystem, should be tested to get better characterization of variance within family or suspected functional group. Problems with the keystone species concept is that it assumes that less well represented species can have disproportionately large impacts on community structure or function. Clearly, dominant species also have a major role in defining the structure of a community and may be equally important as keystone species.

What are the Panel's thoughts on expanded testing of species in sensitive structure/function groups?

It will be important to identify additional species and conceptually it makes sense to prioritize based on functionality in the community or ecosystem (Boutin and Keddy, 1993; Boutin et al., 1995; Boutin and Rogers, 2000). The hard part will be deciding what species should be selected over others when little is known about the relative functionality or importance of each plant species in most ecosystems. Structure and dominance may be a first best guess at relative importance, but trophic relationships should also be acknowledged. It is encouraged that new protocols focus on functional ecology as the selection of species in lieu of strict adherence to taxonomic groups. One advantage of this approach is the ability to offer some insight into how the effects might be mediated at the community or ecosystem level, but this is often difficult to do using taxonomic schemes of selection.

Rather than expanding the number of species tested, the Agency should first determine the optimal number of species required to define the species sensitivity distribution, then determine the optimal species representation. Thus, there is no *a priori* assumption of how many species to test; rather, the extant data base should provide an empirical basis for this determination. It is likely that the types of species tested to generate the species sensitivity distribution will be based on a mixture of taxonomy, physiology, and functionality which may be hard to differentiate when sorting out why particular species are sensitive and others are not. Note that sensitivity is a direct function of the mode of action of the chemical and its relationship to the plant's physiology. Therefore, the species to be tested at Level 1 should be broad enough to encompass all potential modes of action, while Level 3 testing could focus more on species that are known to be sensitive to answer particular, targeted questions.

Question 14 <u>Aquatic and terrestrial multi-species testing</u>

Multi-species testing is proposed at the comprehensive level of assessment (Level 4). Population dynamics and community structure could be affected due to differences in chemical sensitivity among individual species. This may result in

an alteration of plant community structure which subsequently may lead to adverse effects on organisms at higher trophic levels. Multi-species studies provide necessary and invaluable information about changes in population and community dynamics that result from phytotoxic impacts.

How useful are data generated from multi-species/community level studies?

When doing these higher Level tests at Level 3 or 4, it is very important to frame the question being asked. It is not necessary to develop standardized protocols for this, as they will be done for different reasons in the risk assessment progression. These studies are most necessary for examining indirect effects, as direct toxicological effects to single species are conservatively estimated by the Level 1 and 2 assessment. However, the way in which the data will be used in a regulatory context must be clearly spelled out (e.g., will the decision be based on structure or function? How will "adverse effect" be defined, given that most agroecosytems are highly artificial communities (even in the "natural" areas) and that most impacted systems do not "recover" back to the initial state.) Ecosystems are in a constant state of flux and many mesocosm and microcosm studies have shown that these systems do not return to initial conditions once perturbed. These studies also are highly variable, as are natural systems, so questions about how to deal with such variability must be clarified before beginning such studies.

Community studies should not be used at Level 1 of the risk assessment process. This part of the process should be highly conservative; if the species and endpoints are correctly selected and no adverse responses are measured, then there will be no effects to the community. One Panel member expressed concern that single species tests may be too artificial and will not adequately predict the response of organisms that are subject to competition and other stressors in the real environment. Comparison of greenhouse and field data that currently exist should help in determining whether the sensitive endpoints measured at the lower levels of application are over- or underpredictive of field effects. This would be part of the research and development needed to incorporate multispecies analyses into the risk assessment process.

When is multi-species testing appropriate in the proposed design (i.e., how should it be triggered)?

Multi-species testing should be triggered in Levels 3 and 4 or by effects in Level 1 on keystone species when the range of toxic activities could affect several groups or families of plants. The costs and time required to monitor until recovery could be problematic. However, data should exist. Recent efforts to control exotic species in natural communities are expanding rapidly. Standardization will be a problem.

How many trophic levels should be considered in a multi-species test when considering the risk of chemicals to plants?

All trophic levels should be considered. Those selected will depend on the question.

Question 15 **<u>Aquatic and terrestrial post-registration monitoring</u>**
Post-registration monitoring is proposed at the comprehensive level of assessment (Level 4) when adverse effects are anticipated for sensitive species or groups (identified at Level 3). The location and number of monitoring studies will depend on the sensitive species or groups identified and on the types of eco-regions in which they occur. A monitoring study can focus on an indicator species expected to be sensitive, or a multi-species testing design can be introduced to consider the effects on the whole community.

What are the advantages and disadvantages of monitoring studies focused on indicators versus multi-species (communities)?

It was not clear in the documentation provided how ecological monitoring was to be implemented at any level as a preventative or predictive tool. However, it was clearly stated that if monitoring were to be recommended by the regulatory agency, it would be at a higher level (e.g., Level 4) and it would more than likely be predicated on uncertainties identified at earlier levels. It was also understood that it would be a post-registration activity most likely triggered by incidence reporting. With this in mind, the Panel recommends to the Agencies that there are relatively inexpensive and useful mitigative programs to reduce the frequency and occurrence of these incidence reports and/or greatly ease the investigation of such incidences.

The first of these mitigative measures is a program to educate and certify pesticide applicators. In Canada and in most states in the USA, applicators of pesticides must take a two-day course on how to safely handle, store, transport, and use (mix and apply) pesticides. They must pass an examination to ensure that the they can read the labels of the pest control products and do the mathematics required to calculate and formulate the correct tank mixes for delivering the correct recommended applications rates of a chemical or formulated product. Education on these aspects, as well as the importance of prescribed buffer zones and spray equipment, can contribute significantly to reducing the number or frequency of reported incidences of non-target plant damage or impacts on non-target aquatic organisms.

The second mitigative measure is a program that has been established in the UK for licensed applicants of pest control products applied using ground sprayers. It is the legal obligation for the applicant to conduct a local environmental risk assessment for pesticides (LERAP) each time he/she applies pesticides to the environment. Each applicant is provided with guidance and training as to how to complete this assessment; it is easy, does not require much work or time, establishes for each unit of land discrete buffer zones to protect surface waters and sensitive areas, and provides the applicator with a form of due diligence to demonstrate that he/she is in compliance with respect to the application and use of pest control products. It is also a valuable source of information for the regulatory agency in terms of investigation of incidence reports or for investigating what and how much is used, when, where, and by whom.

It is also important to be aware that not all incidences of non-target adverse ecological effects are

the result of off-site migration of pest control products from the site of application. It is possible that some of these incidences could have been caused by non-chemical stressors. Therefore, monitoring programs that are invoked post-registration must be able to discern between chemical (e.g. herbicide) and non-chemical (e.g. water stress; insect damage, disease, etc.).

Monitoring programs must be designed as part of the problem formulation of an ecological risk assessment. The monitoring program must be designed to collect useful information and data that will enable the characterization of risk as it relates to the assessment endpoints (i.e., what it is that must be protected). The design of the monitoring program is, therefore, highly dependent on the goals and objectives of the ecological risk assessment and the magnitude and extent of the problem (i.e., the spatial and temporal scale of the area being affected).

Monitoring programs serve two purposes in a regulatory framework. First, they address uncertainties associated with issues identified at an earlier level (e.g., Level 2 or 3). These usually involve environmental monitoring on a large scale, at multispecies or community level, and likely involve different trophic levels. Use of tools such as remote sensing and GIS methodologies as well as the generation of community indices are appropriate for this level of monitoring. Second, monitoring programs can be directed to answer specific problems that usually involve smaller scales and the use of indicator species and/or specific exposure pathways. Indicator species are often better tools for identifying or monitoring exposure whereas multispecies or communities are often more appropriate for measuring effects. It is not part of the ecological risk assessment paradigm to invoke the use of prescriptive methodologies for monitoring. It is part of the ecological risk assessment framework to design the monitoring program to address the conceptual models that arise from problem formulation.

Thus, the questions being addressed by post-registration monitoring need to be clearly spelled out, as gathering of this type of information can become quite expensive. The general idea is to determine the accuracy of the risk assessment predictions and whether or not mitigation rules are sufficiently protective. The large spatial scale over which some of the herbicides are used and the interactions of other environmental stressors, can make this determination difficult. Two approaches can be used: large scale remote sensing or targeted field studies.

Large scale remote sensing would make use of satellite imagery of the sort that has been described previously in this SAP review. This involves spectral analysis, often in the nonvisual spectrum, to look for changes in pigmentation indicative of plant stress. Patterns of such changes would be overlaid in a GIS with cropping practices and herbicide use patterns. Areas of juxtaposition would be investigated further (on the ground) to attempt to determine cause-and-effect between herbicide use and plant stress. Further work would be needed to determine if previously defined "adverse effects" had occurred (e.g., was plant biodiversity decreased, did community composition either in structure or function change, were individual species affected, etc.).

For aquatic systems, large scale remote sensing may also be applied in a similar fashion, as well as used to ascertain whether algal blooms or die-offs occur in emergent vegetation, etc. As before, cause-and-effect studies would need to be done on the ground. Both of these approaches

beg the question of who would pay for such work and they are not chemical-specific.

Local scale monitoring would be a targeted field study, similar to (albeit on a smaller scale) than those that have been done previously for pre-registration testing. For this approach, it is very important for the questions to be well defined before undertaking such a study, and it must be clearly stated what triggers various levels of management decisions. For these, either indicator species or community studies could be done, the pros and cons of which are discussed below. These would be chemical-specific and designed to determine the efficacy of mitigation efforts.

What criteria are used in the selection of an indicator plant species?

Selection of indicator species would be probably based on prior knowledge and information used to define or describe the situation during problem formulation. Attributes of indicator species that might be desirable include: 1) wide distribution, 2) economically or ecologically important (most reported "ecological" incidences are aquatic in origin; most reported "economic" incidences are terrestrial in origin), 3) sensitive to the chemical of concern, 4) possess a relatively high degree of selectivity, and 5) exhibit consistent symptoms that are easily measured and concentration dependent.

However, if the Agency is concerned about ecological risks, a community analysis may be more appropriate. It is very difficult to extrapolate individual level changes to ecological effects at the community level. Indices of species richness, dominance, and cover can provide a better indication about resiliency of the various plants within the community and whether functionality is maintained even if structure is altered. Indicator species may be most appropriate for cause-and-effects determination or exposure monitoring, through characteristic gross pathologies, biomarker changes, or residue measures. Selection of the appropriate species is based on knowledge of an adequate magnitude response of the exposure marker to the herbicide and the specificity of the response to chemical stress (it wouldn't do to measure a general stress protein, for example, as it could be induced through heat or water stress as well as herbicide exposure). Regardless of the approach used, it is extremely important that exposure be documented so cause-and-effect relationships can be established rather than mere correlations. Plants are continually subjected to environmental stress that can mimic herbicide effects; without clear evidence of exposure there can be no conclusion of causality.

Question 16 **<u>Bioaccumulation</u>**

There is a potential for bioaccumulation of chemicals in non-target plants. Bioaccumulation may be one indicator of hazard and more importantly an indication of the extent of uptake and translocation of chemicals in plants. Chemicals that bioaccumulate in plants may also have implications for herbivorous wildlife species. The Agencies are less certain on whether to assess the effects of bioaccumulation in the determination of overall risk to non-target plants.