SAP Report No. 2001-06

REPORT

FIFRA Scientific Advisory Panel Meeting, March 13-16, 2001, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Probabilistic Models and Methodologies: Advancing the Ecological Risk Assessment Process in the EPA Office of Pesticide Programs

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.

SAP Report No. 2001-06

REPORT

FIFRA Scientific Advisory Panel Meeting, March 13-16, 2001, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Probabilistic Models and Methodologies: Advancing the Ecological Risk Assessment Process in the EPA Office of Pesticide Programs

Mr. Paul Lewis
Designated Federal Official
FIFRA Scientific Advisory Panel
Date:

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

Stephen M. Roberts, Ph.D.
Session Co-chair
FIFRA Scientific Advisory Panel
Date:

CONTENTS

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
March 13-16, 2001

Probabilistic Models and Methodologies: Advancing the Ecological Risk Assessment Process in the EPA, Office of Pesticide Programs

PARTICIPANTS

FIFRA SAP Chair
Ronald J. Kendall, Ph.D., Professor and Chairman, Department of Environmental Toxicology and Director, The Institute of Environmental and Human Health, Texas Tech University and Texas Tech University Health Science Center, Lubbock, TX

FIFRA Scientific Advisory Panel
Fumio Matsumura, Ph.D., Professor, Institute of Toxicology and Environmental Health University of California at Davis, Davis, CA
Christopher Portier, Ph.D., Acting Associate Director, National Toxicology Program National Institute of Environmental Health Sciences, Research Triangle Park, NC
Stephen Roberts, Ph.D., Director, Center for Environmental and Human Toxicology, University of Florida, Gainesville, FL

FQPA Science Review Board Members
William Adams, Ph.D., Kennecott Utah Copper Corporation, Magna, UT
Louis Best, Ph.D., Department of Animal Ecology, Iowa State University, Ames, Iowa
George Cobb, Ph.D., Texas Tech University, Institute of Environmental Health, Lubbock, TX
Peter Delorme, Ph.D., Environmental Assessment Division, PMRA, Health Canada Ottawa, ON, Canada
Philip Dixon, Ph.D., Department of Statistics, Iowa State University, Ames, IA
Paul Eslinger, Ph.D., Pacific Northwest National Laboratory, Richland, WA
M. Elizabeth Halloran MD MPH Dsc, Department of Biostatistics, Emory University, Atlanta, GA
Kirk Hatfield, Ph.D., University of Florida, Gainesville, FL
Andrew Hart, Ph.D., Wildlife Ecotoxicology Unit, Central Science Laboratory, Sand Hutton, York, United Kingdom
Steven Heeringa, Ph.D, Institute of Social Research, University of Michigan, Ann Arbor, MI
Charles Henny, Ph.D., United States Geological Survey, Corvallis, OR
Chad T. Jafvert, Ph.D., Professor and Area Head, Environmental and Hydraulic Engineering

Area, School of Civil Engineering, Purdue University, West Lafayette IN
Thomas W. LaPoint, Ph.D., Department of Biological Sciences and Institute of Applied Sciences, University of North Texas, Denton, TX
Pierre Mineau, Ph.D., National Wildlife Research Center, Canadian Wildlife Service Hull Quebec, Canada
Michael Newman, Ph.D., Office of Dean of Graduate Studies, Virginia Institute of Marine Sciences, Gloucester Point, VA

PUBLIC COMMENTERS

Oral statements were made by:
Bill Williams, Ph.D. on behalf of the American Crop Protection Association
David Fischer, Ph.D. on behalf of the American Crop Protection Association
Iain Kelly, Ph.D. on behalf of the American Crop Protection Association
Paul Hendley, Ph.D. on behalf of the American Crop Protection Association
Jeff Giddings, Ph.D. on behalf of the American Crop Protection Association

No written statements 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 probabilistic models and methodologies: advancing the ecological risk assessment process in the EPA, Office of Pesticide Programs. Advance notice of the meeting was published in the *Federal Register* on February 14, 2001. The review was conducted in an open Panel meeting held in Arlington, Virginia, on March 13-16, 2001. The meeting was chaired by Ronald J. Kendall, Ph.D. Mr. Paul Lewis served as the Designated Federal Official.

In April 2000, the EPA, Office of Pesticide Programs (OPP), Environmental Fate and Effects Division (EFED), met with the SAP to present a progress report on its initiative to revise the ecological assessment process for pesticides. This included an update regarding the progress of the initiative and corresponding Panel recommendations on the approach. A key component of the April 2000 meeting was an overview of the conceptual risk assessment model proposed by OPP. Since the meeting, OPP has conducted a generic case study which incorporated many of the comments and recommendations by the SAP. The purpose of this 4-day meeting is to review the generic case study with the SAP.

The meeting was divided into two parts. The first two days focused on review of a probabilistic model to assess acute lethal risks to birds. Mr. Edward Fite (EPA, Office of Pesticide Programs) provided an introduction to the topic. Timothy Barry, Sc.D. (Office of Economy and Environment, EPA) provided an overview of the EPA, Office of Pesticide Programs' Pilot Avian Risk Model; Edward Odenkirchen, Ph.D. (Office of Pesticide Programs,

EPA) and Edward Fite, M.S. (Office of Pesticide Programs, EPA) presented a probabilistic model and process to assess acute lethal risks to birds. The final two days of the meeting concerned a probabilistic model and process to assess risks to aquatic organisms. Kathryn Gallagher Ph.D. (Office of Pesticide Programs, EPA), provided an overview of the aquatic assessment, Timothy Barry, Sc.D. (Office of Economy and Environment, EPA) presented an overview of the aquatic risk assessment model, James Lin, Ph.D. (Office of Pesticide Programs, EPA) summarized the aquatic exposure assessment and Les Touart, Ph.D. (Office of Pesticide Programs, EPA) presented methods and approaches to conducting an aquatic effects assessment.

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

Probabilistic Models and Methodologies: Advancing the Ecological Risk Assessment Process in the EPA, Office of Pesticide Programs

A Probabilistic Model to Assess Acute Lethal Risks to Birds

CHARGE

1. Focal Species Selection Goals: The goals for selection of bird species serving as the focus of the risk assessment were to

> (1) advance the assessment beyond consideration of "generic" bird types so as to consider appropriate biological conditions associated with the treated environments, and
> (2) identify the types of species potentially at greatest risk from ChemX exposure at the corn and alfalfa use sites.

Under the current EFED status of probabilistic risk assessment development, the use of focal species in an assessment is limited. The likely lack of species-specific toxicity data engenders considerable uncertainty regarding the prediction of the magnitude of mortality in any single bird species. Rather, the use of focal species is targeted to represent a myriad of potential species of similar biological/behavioral characteristics, yet retain some specificity as to the type of organisms using a treated area.

- What is the Panel's opinion regarding this approach?
- What are the Panel's recommendations regarding alternative approaches that should be investigated?
- Drawing upon your experiences and knowledge of avian foraging strategies, habitat use, and other interactions with the agroenvironments incorporated into the assessment, what recommendations can the Panel make for alternative or additional focal species for the crop/region combinations investigated in the assessment?

2. Frequency of Birds in Treated Fields use in the Model: EFED recognizes that additional research on quantifying exposure of bird species in agroenvironments will be critical to the advancement of the probabilistic risk assessment approach. At this juncture, EFED is mindful of the severe limits of existing avian census data for establishing such exposure estimates. The present method for considering the avian census data has been designed not to over represent the census data to the point that sightings of birds on or off a treated field is considered commensurate with proportional feeding on and off the field in a given day. Instead, the risk assessment uses the data to determine the likelihood that a bird will be on a treated field in a given time step, based on past field study history of sightings for that species.

- What are the Panel's thoughts regarding the use of avian census data in the model and was it used appropriately?
- What are the Panel's suggestions regarding alternative approaches to using the data? Please discuss any advantages and disadvantages to these alternative approaches.

3. Frequency of Birds in Treated Fields, Setting Parameters: EFED elected to establish minimally biased truncated exponential distributions for this parameter for each focal species.

- Upon looking at the available field study data (see spreadsheets included in SAP package), what are the Panel's thoughts on these selections?
- What alternative approaches for these distributions would be appropriate for the data sets available?

4. Consideration of Drinking Water Source Selection: Because of the paucity of data for drinking water source selection in birds of agroenvironments, EFED has identified investigation of this behavior as an area meriting further research. However, in the interim, EFED has made an assumption that drinking water selection is opportunistic and that use of on-field water sources is linked with bird presence on the field during any particular time step (see questions on frequency of birds using treated fields).

- What are the Panel's views on this interim procedure?
- What recommendations can the Panel make for future alternative approaches, considering the data currently available to EFED?

5. Puddle Persistence: EFED recognizes that on-field puddles may be more persistent than the half-day assumption incorporated in the model. However, EFED does not currently have a way of modeling puddle duration in a field.

- What are the Panel's recommendations regarding data sources and/or modeling approaches to establish the frequency of occurrence, dimensions, and duration of puddles in agroenvironments? (Rhetorical: When does a puddle become a pond?)

6. Concentrations in Drinking Water:

A. In the absence of a more rigorous model for pesticide residue in dew, EFED has used a simple two-compartment partitioning model. In the Case Study document, EFED has discussed some critical limitations to this model.

- What model modifications can the Panel suggest for improving the estimation of pesticide residues in dew, keeping in mind the limitations of the current registration data set requirements?
- What, if any, data sources regarding dew measurement or modeling should be considered to strengthen EFED's modeling needs?

B. EFED used an instantaneous maximum approach from direct application for modeling pesticides in puddles present in a treated field on the day of application. This approach, based on experiments with buried buckets of water, does not consider partitioning kinetics with field soils.

- What does the Panel suggest for improving this approach?

C. Pesticide residues in puddles appearing on a treated field on the day after pesticide application were estimated using output from PRZM runs. The operative assumption in this approach is that water puddling on the surface of the field would be equivalent in pesticide concentration to the water potentially running off the field from a given precipitation event. EFED recognizes a number of limitations to this approach, not the least of which is the sensitivity of the PRZM model output to application date in terms of number and magnitude of run-off event pesticide loadings following the application date.

- What are the Panel's suggestions for adapting existing tools to this task and improving the modeling of day-after treatment puddle residues?

7. Residues in Vegetation Food Items: In the Case Study, EFED chose to base estimates of pesticide residues on the data provided in Fletcher et al. (1994). EFED also had data on actual ChemX residues after treatment, but these data were severely limited in sample number and in a very limited number of field study areas. EFED based the decision to use Fletcher et al. (1994) on a desire to utilize a more robust, albeit non-chemical specific, data set to establish initial pesticide residues.

- What is the Panel's opinion regarding EFED's approach, which relies on larger more generalized data sets versus focusing on limited single-chemical data sets for estimating initial field concentrations of pesticide residues
-

8. Residue Clearance in Focal Species: EFED has included a residue retention factor in the exposure model. This factor was intended to account for carryover of a proportion of a time step's body burden to the next exposure time step, allowing for some consideration of cumulative exposure.

- What are the Panel's thoughts on this approach?
- Would the Panel recommend approaches to account for uncertainty associated with extrapolating poultry metabolism data (the origin of the residue retention factor used in the Case Study) to other bird species?
- Please comment on any need for pharmacokinetic studies to improve this assessment approach. What are the Panel's recommendations on the scope of such studies (i.e., appropriate species, number of species, study design, and endpoints)?

9. Selection of Exposure Time Steps: The Case Study divides each exposure day into two time steps. There have been questions regarding the effect of exposure time step and feeding duration upon the outcome of the model. To investigate this, EFED used a mass balance equation with two compartments, to determine the effect of rate of food consumption combined with clearance rate, on total body burden (ingested + residual) of pesticide. The results of this exercise led EFED to conclude that duration of the exposure window within a time step had minimal effect on maximum instantaneous body burden of ChemX.

- Given the lack of detailed pharmacokinetic data for ChemX in birds, what other analyses of this issue could be made?

10. Selection of Acute Toxicity Standard: In the Case Study, EFED used the acute single oral dose studies (LD_{50}) as the basis for characterizing effects in preference to the acute dietary exposure studies. Both of these study designs have limitations for estimating the risk to wild avian species exposed to pesticides in the environment. Both studies have a fixed exposure period, not allowing for the differences in response of individuals to different durations of exposure. For the acute oral study, the dose administered in a single dose all at one time does not mimic wild birds' exposure. Also, for exposure through different environmental matrices, it does not account for the effect of the matrices on the absorption rate of the chemical into the animal.

This latter criticism also applies to the dietary test for other food matrices consumed in the wild. For the dietary test, the endpoint is reported as the concentration mixed with food that produces a response rather than as the dose ingested. There are a number of study limitations that render conversion of dietary concentration to dose problematic (e.g., food spillage, and quantifying food intake for individuals). The interpretation of this test is also confounded because the response of birds is not only a function of the intrinsic toxicity of the pesticide, but also the willingness of the birds to consume treated food.

More importantly, there is evidence for some compounds that the laboratory derived LC_{50} values are poor predictors of effects in the field.

- What are the Panel's thoughts and recommendations regarding alternative approaches using LC_{50} data that allow estimation of dose-response relationships (critical to prediction of magnitude of effects)?

11. Defining the Distribution of Species Sensitivity: EFED has selected a combination of methods to establish the distribution of species sensitivity to ChemX, from which representative points, low, medium, and high sensitivity are used to characterize the uncertainty regarding response of any particular species to ChemX exposure. The first step in the method involves normalizing all available toxicity data to a constant body weight using weighting factors established for specific and generic chemicals. The mean and standard deviation of these normalized values is taken. Then the method of Aldenberg and Slob (1993) is used to estimate the 5^{th} and 95^{th} percentiles of a log-logistic distribution with mean and standard deviation as

defined above. Finally these point estimates of sensitivity are readjusted to a focal species body weight.

- What are the Panel's views on this method to estimate the distribution of sensitivity of focal species when species-specific toxicity data are not available?
- How would the Panel estimate the confidence interval surrounding the above estimates of the 5^{th} and 95^{th} percentiles?

12. Other Factors Affecting Sensitivity: By relying on laboratory LD_{50} data, we have not accounted for a variety of physiological and environmental factors that may modify sensitivity (e.g., age, nutritional status, temperature, etc.) . EFED has concerns that not accounting for a number of these variables may lead to an underestimation of risk.

- What guidance can the Panel offer that would allow, in the absence of chemical specific data, for a consideration of these physiological and environmental factors?

DETAILED RESPONSE TO THE CHARGE

The specific issues to be addressed by the Panel are keyed to the Agency's background document "Transmittal of Review Documents for the March 13-16, SAP Meeting," dated February 16, 2001, and are presented as follows:

General Comments

The FIFRA Scientific Advisory Panel congratulates the Agency on the effort made to conduct probabilistic risk assessment of pesticide effects in ecosystems. The approach has progressed greatly from paradigms discussed at initial Ecological Committee on FIFRA Risk Assessment Methods (ECOFRAM) meetings. The intricacy of the models was surprisingly good given the time interval that the Agency had to complete this task. As with any scientific endeavor, there is still work to be done. These comments are provided to facilitate future efforts by the Agency.

A Probabilistic Model to Assess Acute Lethal Risks to Birds

1. Focal Species Selection Goals: The goals for selection of bird species serving as the focus of the risk assessment were to

> **(1) advance the assessment beyond consideration of "generic" bird types so as to consider appropriate biological conditions associated with the treated environments, and**
> **(2) identify the types of species potentially at greatest risk from ChemX exposure at**

the corn and alfalfa use.

Under the current EFED status of probabilistic risk assessment development, the use of focal species in an assessment is limited. The likely lack of species-specific toxicity data engenders considerable uncertainty regarding the prediction of the magnitude of mortality in any single bird species. Rather, the use of focal species is targeted to represent a myriad of potential species of similar biological/behavioral characteristics, yet retain some specificity as to the type of organisms using a treated area.

- **What is the Panel's opinion regarding this approach?**
- **What are the Panel's recommendations regarding alternative approaches that should be investigated?**
- **Drawing upon your experiences and knowledge of avian foraging strategies, habitat use, and other interactions with the agroenvironments incorporated into the assessment, what recommendations can the Panel make for alternative or additional focal species for the crop/region combinations investigated in the assessment?**

Selection of Focal Species

The Panel agreed that the use of focal species to represent potential species of similar biological/behavioral characteristics is reasonable. Detailed analysis of all species potentially at risk from each individual pesticide to be reviewed by the Agency is impractical, and use of focal species, thoughtfully selected, can probably adequately represent the potential exposure of the avian community at risk. The uncertainty relative to species-specific toxicity is likely to be greater than the uncertainty relative to species-specific habitat-use patterns, feeding ecology, and other life history traits.

The list of types of information that could serve to identify focal species (i.e., direct toxicological evidence, known occurrence in treated crops, known mortalities, life history characteristics) seems reasonable. The relative importance placed on the various selection criteria, however, could influence the species selected. While the Agency's background document provided a general list of criteria used to identify focal species, it did not indicate the relative importance placed on various selection criteria. The more transparent the selection process, the more readily it can be reviewed and improved.

In the broader sense of overall risk assessment, criteria for selecting focal species depends on what question the risk assessment is intended to answer, i.e., on what the risk manager wants to know. If the intent is to develop statements about the predicted mortality of species that are selected to be 'worst case' with regard to exposure and known history of field kills, then the factors considered in selecting the focal species would be appropriate. If information is needed about the percentage of species that will experience given levels of mortality, then the population of species considered needs to be defined and the focal species need to be chosen so as to be statistically representative of them.

Not specifically mentioned is information on nest-site location. Whether the species nests within treated fields or in habitats surrounding treated fields could influence potential risk, particularly for nestlings. This is important because toxicity may differ between adult and young birds. Knowledge of nest-site location also could provide insight into the probable use of treated fields by birds (Mortensen et al, 1996). Birds nesting in edge habitats likely visit crop fields primarily to feed, whereas birds that nest in crop fields use the field for other activities as well (e.g., incubation and brooding, roosting).

In addition, the Agency did not discuss the seasonal use of treated fields by birds. As the corn crop develops, the fields are transformed from barren, sparsely vegetated habitat into dense plant cover that can attain heights over 2 m. This has a dramatic effect on the bird species using such fields and is relevant to risk assessment depending upon when the pesticide is applied. Species that are most abundant in crop fields early in the breeding season (e.g., horned lark, killdeer) eventually are replaced by forest edge species (e.g., black-capped chickadee, indigo bunting) (Best, 2001). Also, the timing of pesticide application relative to mowing of alfalfa fields is important. Bird use of alfalfa fields differs dramatically before and after mowing, and recolonization of alfalfa fields by birds after mowing depends upon crop regrowth (Frawley and Best, 1991). Some species will recolonize such fields (e.g., dickcissel), but others will not (e.g., red-winged blackbird).

Critical to the selection of focal species is the adequacy of field data on bird use of cropland. A substantial amount of information on bird abundances in a variety of crops and geographical regions was gathered during the avian field studies required by EPA. Many of these studies followed similar protocols, thus facilitating comparison of study results. Unfortunately, most of these data are not readily accessible. This argues for the development of a data depository or some other means to make the information more generally available to the public. The published information on bird abundance, habitat-use patterns, breeding ecology, and foraging ecology available for birds in midwestern cornfields is not likely to be replicated for other crops and regions.

When using bird survey information to decide on focal species, it is important to distinguish bird use of crop fields from bird use of edge habitats adjacent to crop fields. Some bird species may be abundant in edge habitats but rarely visit crop fields, and the distances that edge species venture into crop fields also differ (Best et al. 1990; Schiavone and Best, unpublished, data).

Use of mortality incidents as a criterion for selecting focal species should be used with caution. Habitat-use patterns and conspicuousness differ among species and would influence their ability to be found. Large birds are more conspicuous than small birds, and field resident species would more likely be found than species whose primary habitat is off site. Also, greatest reliance should be placed on mortality incidents where the cause of death can be confirmed by residue analysis (Kendall et al., 1992).

Considering all of these parameters, it would be prudent to add species that feed in the crop canopy. All those species selected feed on the ground (page 6 of the Agency's background document) and while being more susceptible to granular applications, may be less susceptible to foliar applications. Birds that consume some reasonable amount of herbage and that use the crop lands under assessment should also be evaluated (Table 4, p. 14 of the Agency's background document).

Focal Species for Corn and Alfalfa

The Panel specifically addressed focal species for corn and alfalfa:

Focal species for corn - the selection of meadowlarks as a focal species for corn could be questioned. They rarely use cornfields (Best et al., 1990) and have only been documented to nest in no-till fields (Best, 1986). One could also question why the American robin was not included. This species regularly uses cornfields, dead birds have been found in treated fields, and the robin represents a dietary route of exposure (vermivore) not included in the assessment.

Focal species selected for alfalfa - Patterson and Best (1996) was used as a reference for focal species selection for alfalfa. This is based on a misreading of the paper. The Conservation Reserve Program (CRP) fields used in that study were predominantly smooth brome; some did have a mixed seeding that included alfalfa. Why wasn't the red-winged blackbird included among the focal species for alfalfa? It is one of the most abundant bird species breeding in midwestern alfalfa fields before mowing (Frawley and Best, 1991). Here the timing of the pesticide application relative to mowing is critical because the abundance of bird species in alfalfa fields before mowing differ substantially from those immediately after mowing. The red-wing blackbird also would seem to be a good choice because of the availability of toxicity data.

The use of the mallard to represent gorge feeding on alfalfa is questionable. During the breeding season, this species is primarily a granivore (unlike geese). Thus, the assumption that half of the daily food requirement consisted of contaminated alfalfa is probably false. Also, it is doubtful that these birds would be flocking at the time of year when ChemX would be applied (nesting starts in mid April in Iowa). This indicates the importance of understanding the temporal relationship between breeding phenology and pesticide application.

Generic species versus actual species

Panel members commented on the relative advantages of using generic species or actual species in the risk assessment. One Panel member concluded that using generic species could be simpler, but if one reaches the point where use of detailed field data on exposure factors (such as proportion of time spent in crop) is needed, then the assessment inevitably will become less generic, and relevant primarily to the species for which those data are obtained. This is consistent with the ECOFRAM approach, wherein generic species could be used at lower levels of refinement and be replaced with focal ones at higher levels.

The Agency had outlined the idea of developing a library of exposure scenarios for species/crops/regions. One reason for suggesting a switch to generic species is the resource cost of characterizing a large number of actual species scenarios. However, if we can ensure that the assessment model quantifies the uncertainty in the scenarios and accounts for its effect on the risk estimate, then the scenarios need not be very precise and could be based on existing information. The resulting risk estimates will have large confidence intervals but that is appropriate as it is a reflection of the true state of our knowledge.

Risk estimates based on generic species would actually have similarly high uncertainty due to the variable relationship between exposure factors for the generic species and actual species. In principle, this uncertainty should also be quantified and this could be done using the same set of information on actual species.

If done properly (i.e., in a way which accurately reflects our uncertainty about the risk), both approaches require similar amounts of effort. However, the approach based on actual species has the advantage that the risk estimates can be expressed in relation to actual rather than generic species, which may have more utility for use by the risk manager.

A key issue here is whether there is an attempt for a full account of the uncertainty affecting exposure estimates. Concern was expressed that the resulting risk estimates would have extremely wide confidence bounds, possibly ranging from zero to 100% mortality for many pesticides. However, if our estimates are that uncertain, the Agency would want to know this. Also, even if such an analysis did not provide an adequate basis for screening out some pesticides as having less risk, it could be used to identify the key factors that should be addressed in refined assessment. Furthermore, over time, the refined assessments would generate more precise data for the exposure scenarios which (subject to data-sharing arrangements) could be used to progressively reduce the uncertainty present in the base-level assessment.

2. Frequency of Birds in Treated Fields use in the Model: EFED recognizes that additional research on quantifying exposure of bird species in agroenvironments will be critical to the advancement of the probabilistic risk assessment approach. At this juncture, EFED is mindful of the severe limits of existing avian census data for establishing such exposure estimates. The present method for considering the avian census data has been designed not to over represent the census data to the point that sightings of birds on or off a treated field is considered commensurate with proportional feeding on and off the field in a given day. Instead, the risk assessment uses the data to determine the likelihood that a bird will be on a treated field in a given time step, based on past field study history of sightings for that species.

- **What are the Panel's thoughts regarding the use of avian census data in the model and was it used appropriately?**
- **What are the Panel's suggestions regarding alternative approaches to using the data? Please discuss any advantages and disadvantages to these alternative**

approaches.

The Panel agreed that census data are more likely to reflect abundance of birds than proportion of food that individuals obtain from a given crop. The presence of a bird does not necessitate foraging. Differences in observability could also bias the time in crop data. The solution to this problem is radiotracking, and in the case of nestling exposure, radiotracking of adults with video monitoring of feeding activity in the nest. Both techniques have been successfully deployed in artificial and natural nests. This would also set the stage for higher tier assessments where model outcomes must be verified in the field.

The current approach uses time on treated fields as a proxy for the percent of daily food and water units that contain residues. This model assumes each food unit eaten from the field is contaminated and therefore includes no stochastic variability in the pesticide load per food item. The proportion of diet allocated to each food source is fixed for a fourteen day model period. Assignment of individuals to treated field food sources is all or nothing for modeled time intervals. One option would be to allow the stochastic model to simulate daily feeding behavior in which each food source is derived in part from treated fields and the balance from untreated areas.

It would seem that the use of on/off field data for 12 hour steps may misrepresent activities and thus exposure potential. It is difficult to conceive that allowing more frequent choices for foraging or not does not alter the ingestion rate, especially in light of the underlying foraging distributions. The Agency used unimodal distributions of foraging time on treated fields, but radio-tracking studies of individual birds in the UK show that for some species, time in crop is bimodally distributed.

The Panel believed that the Agency should use caution when evaluating field data that represent percentage activity on different sites. Such data reflect the average behavior of birds on each site. Distributions fitted are therefore distributions of average behavior between sites. However, the Agency is using it in the model as a distribution of individual behavior. In reality, the distribution of individual behavior within site could be very different from the distribution of average behavior between sites. For example, the assumption of no serial correlation (p. 16) between sequential foraging events is clearly unrealistic, especially for territorial species, and will lead to significant under-estimation of risk for a proportion of individuals.

The Panel agreed that the Agency should avoid assuming that time in crop equals food from crop, because it is likely that feeding rates in and out of crop will often be different. The Agency's response was to use the field observations to assign 12 hour periods to crop/non-crop. However, (a) the Agency's estimate of food from crop is still entirely driven by "time in field" observations, so it doesn't really address the possibility of feeding rates differing and (b) the Agency's construct is adding unrealism, as birds do not really behave as described in the model. Also, the AgDrift Task Force has data demonstrating drift from fields and this data should be incorporated into a concentration density surrounding a field. An alternative might be to add the uncertainty about feeding rate differences explicitly into the model, using expert judgement to

assign bounds, and examine their effect on the confidence limits on risk. Foraging in edge areas with marginal pesticide exposure could also be added to the assessment. If either effect is large, it indicates that the issue is critical to the assessment and one will simply have to find ways of quantifying it more precisely.

Finally, all of the uncertainty surrounding this critical behavior highlights an area where serious attention is needed. Data in Table 5 of the Agency's background document represent a small amount of the Avian Census data available to the Agency. The paucity of data used in this assessment demonstrates a serious data usage problem that arises from issues of proprietary data among registrants. In the case of avian census and other data sets, more data are needed in PRAs, and these data exist. If proprietary issues cannot be resolved, then each registrant will be in the position of submitting data packets containing census and radio tracking data with sufficient statistical power to describe foraging within the crop systems in which risk is being evaluated. To generate such detailed data for each focal species in each crop for each pesticide evaluated, would *de facto* require a return to full scale field studies albeit separated into several smaller parts that need not necessarily be conducted simultaneously.

3. Frequency of Birds in Treated Fields, Setting Parameters: EFED elected to establish minimally biased truncated exponential distributions for this parameter for each focal species.

- **Upon looking at the available field study data (see spreadsheets included in SAP package), what are the Panel's thoughts on these selections?**
- **What alternative approaches for these distributions would be appropriate for the data sets available?**

Parameterization of this activity is critical to the PRA process and the Panel commends the Agency for the effort to produce a reasonable estimate with little data. There are several areas where this process could be improved. The first is to clarify the terminology used in describing the PRA process. Also, the Agency must ensure that the procedures developed initially are sufficiently robust to perform well in future PRAs for other pesticides. These observations apply to other aspects of the risk assessment as well.

One example of terminology that should be clarified can be found in the distribution used for avian frequency in treated fields. This distribution is a truncated exponential distribution when the mean is less than 0.5 and a reversed truncated exponential distribution (i.e. a truncated exponential distribution on 1-x) when the mean is larger than 0.5. This distinction is not clear in the Agency's background document.

The truncated exponential distribution has an explicit functional relationship of the variance to the mean of the distribution. Beta distributions for the in-field probabilities would permit separate simulation of mean and variance. The Agency is encouraged to examine model scenarios in which the probabilities of time in field follow a beta distribution with differing means and

variance. In particular, it would be useful to evaluate at least two beta distributions, one with significant mass near zero and 1.0, and a second in which the majority of the distributional mass is in the middle of the distributional range. These two limited alternatives should demonstrate the sensitivity of the model results to assumptions about the distribution of the proportion of treated foods consumed.

In discussions that occurred regarding this topic, an important question arose: Is fitting a distribution to these data worthwhile? The Panel agreed that modeling available data sets is valuable if the chosen assumptions and distributions make biological sense. If the studied fields were a random sample from all possible fields in the target area, then the frequency distribution of the observations is the non-parametric maximum likelihood estimator of the population density function. If the population fits the postulated model (e.g., truncated exponential), then the estimates from the sample data are good estimates of the population parameters. However, the eight studies are not a random sample, so the validity of the frequency distribution cannot be established from statistical properties alone.

There is a continuum of conceptual models: at one extreme, the eight observations are a sample from the target population, even if they come from different data collection methods, different types of edges, and different regions of the Midwest. In this view, all observations provide information about the population of interest (an arbitrary field in the Midwest). At the other extreme, each observation is a unique case. The variability between the observations can be explained by each observations' unique characteristics. If each field is unique, it only provides information about itself. The data provide no information about other fields. Reality is likely to be in between the two extremes. Most fields may provide information about the population of interest, but some fields may not because they represent some other biological setting. The Panel suggests that biological information about each study be used to decide whether or not it describes the population of interest. Those observations that do not represent the population of interest should be deleted before fitting the probability distributions.

To this end, it would be useful for a collaborative effort of biology experts and statisticians to review the studies to determine to what extent the studies are comparable (or exchangeable in the Bayesian sense).

When combining the results of several studies, once the decision has been made that it is biologically meaningful, it would be appropriate to use a hierarchical model, possibly Empirical Bayes or a full Bayesian approach. This would give an estimate of the overall mean, incorporating the variability of the individual studies into the estimation. A nonhierarchical approach would be to take the weighted average of the individual estimates, weighted by the inverse variances of the individual studies. However, this would likely underestimate the variability considerably. Therefore, the Panel does not recommend this simpler approach. There is a large literature on hierarchical models, and in particular meta-analysis (a method of combining results of independent studies), that the Agency could draw on.

Again, drawing on the biology, all species are unlikely to have the same form of distribution for frequency in the treated fields, as indicated by United Kingdom radiotracking studies that found quite different frequency distributions following unimodal and bimodal functions (Crocker, et al., 2001). Therefore, it might be useful to try other standard distributions for the selected species.

Finally the Panel believed that field validation of model output will be critical when PRAs progress past this level. Such validation should be planned into the assessment process.

4. Consideration of Drinking Water Source Selection: Because of the paucity of data for drinking water source selection in birds of agroenvironments, EFED has identified investigation of this behavior as an area meriting further research. However, in the interim, EFED has made an assumption that drinking water selection is opportunistic and that use of on-field water sources is linked with bird presence on the field during any particular time step (see questions on frequency of birds using treated fields).

- **What are the Panel's views on this interim procedure?**
- **What recommendations can the Panel make for future alternative approaches, considering the data currently available to EFED?**

The Panel believed that there is little information in the literature regarding the selection of water sources by birds. A brief search by one Panel member revealed few scientific papers which mention drinking, and only one was really pertinent – it showed that choice of drinking sources by sandgrouse is affected by a preference for open ground (to aid detection of approaching predators). However, other species (especially smaller ones) might have an opposite tendency, and prefer to use water sources under vegetational cover.

Seed-eating birds in particular are often thought to be attracted to puddles on fields, and this is used when choosing locations to catch birds in field studies. However, in one Panel member's experience, some passerines (e.g., horned larks) are not attracted by water baths. It was suggested that the species most represented by the Agency's model are waterfowl, which are frequently observed drinking.

Variation in sources of water used by birds is likely to arise from environmental factors as well as differences among species. Geographic and climatic variation are important both to the presence of sources and birds' requirements and behavior. Also, water intake might be greater in the afternoon when temperatures are higher and there is more heat stress.

There was considerable uncertainty about the extent to which birds take dew. The Panel, and other ornithologists they consulted, were doubtful whether many species would take significant quantities of dew. One thought that columbiformes, with the exception of pigeons, cannot normally ingest liquids.
On the other hand, birds may be able to pick up droplets of dew using surface tension.

The Panel considered that birds are less likely to take dew if reliable standing water sources were available off field. Furthermore, not all dew on-field is on the crop, and the concentration of pesticides in other dew sources may be different. There may, therefore, be a need to consider choice between different sources of dew within the field. Finally, there was uncertainty about whether dew would remain available each day for as long as is assumed in the Agency model.

There was concern that birds may take spray droplets which collect in leaf whorls. It was reported that some bird kills in Germany had been found to be caused by this means (Hommes et al., 1990). One Panel member was aware of re-registration data submitted to the Agency for an organophosphate pesticide, giving measured residue values for pesticide granules occurring in leaf whorls. This would present a risk via drinking if birds took dew or rainwater that subsequently collected in the whorls.

There are other potential sources of drinking water that could contain pesticide including small streams, ditches, and ponds near the point of application. Water in these sources may have lower concentrations than field puddles, yet much more than off-site sources. These resources are unlikely to be important for ChemX but might be for pesticides with lower field decay rates. In such cases, drinking water sources of longer duration may make a significant contribution to exposure.

Suitability of current approach

Overall, the Agency's approach (of assuming birds will take water opportunistically if it is present) seems reasonable, given the lack of good information on selection of water sources by birds.

The linkage between time on the field and the amount of water consumed during that time period is critical and requires further investigation. In addition to the amount of water actually consumed by various species, data on bird drinking habits are essential. The assumptions used in the model are critical to the amount of risk that is actually attributed to waterborne exposure. At present, there appears to be considerable uncertainty as to the extent of exposure that actually occurs under field conditions. In cases where the concentration of the test substance in the

water contributes significantly to potential risk, it would be advisable to verify the exposure.

Recommendations for future approaches

The results suggest that exposure via drinking water can make a significant difference to the risk of mortality and therefore deserves further investigation.

The three specific scenarios modeled (dew vs. dew plus two types of puddles) gave similar results. However, as concentrations in these three sources differ markedly, the choice of scenario might have mattered more if puddle persistence had been allowed to vary over periods longer than half a day, or if different assumptions had been made about birds' use of dew. Therefore, perhaps priority should be given to obtaining better information on puddle persistence, on concentrations in dew and puddles, and on which bird species take dew.

Depending on the outcome of these investigations (e.g., if they indicate a potential for water consumption to contribute >10% of total exposure), it might be necessary to study how birds select water sources in more detail. The data that the Agency typically has available are not sufficient to provide further resolution to this problem. Careful field telemetry studies combined with laboratory bird behavior studies could provide the needed data.

It may be worth expanding the model to consider other water sources adjacent to the treated field such as streams, ditches, and ponds. In effect, the model may already consider these as sources of drinking water - since the values calculated with PRZM are run-off concentrations anyway. However, there are many ways to approach this beyond PRZM, e.g., the farm pond approach used in the aquatic model (EXAMS). This may be an issue that should be reserved for higher-tier assessments, particularly for pesticides which are relatively persistent in water.

5. Puddle Persistence: EFED recognizes that on-field puddles may be more persistent than the half-day assumption incorporated in the model. However, EFED does not currently have a way of modeling puddle duration in a field.

- **What are the Panel's recommendations regarding data sources and/or modeling approaches to establish the frequency of occurrence, dimensions, and duration of puddles in agroenvironments? (Rhetorical: When does a puddle become a pond?**)

The Panel agreed with the Agency that the assumption of puddles existing for a half-day may underestimate the total puddle duration for many

agroenvironments. There would be both inter- and intra-year weather fluctuations that are uncontrollable and may be capable of producing great variance in model output, but capturing such variability is essential for a realistic PRA process. Specifically, the occurrence and lifespan of standing water will be related to rainfall events, solar irradiance, temperature, humidity, topography, elevation of water table, soil properties, field cover, and efficiency and occurrence of field tiles. Given the parameters involved, it may be too cumbersome at a level two refinement to get much more complicated than the Agency's currently presented approach. Perhaps given the selection of high percentile rainfall events, puddles could be allowed to persist for longer durations.

As models are refined at higher levels within the PRA, some simple approaches to modeling puddle size and duration may provide improvements over the current half-day assumption. Data are available from University extension stations and State agricultural extension services regarding the extent of evaporation that occurs within a given county on a monthly basis. Since water is lost from puddles through evaporation and infiltration, pan-evaporation data provide a reasonable estimate of the time it will take for a given amount of water (e.g., one inch) to evaporate under a given set of conditions. For example, in Salt Lake County in July, the evaporation rate is one inch per day. In more humid environments (Kansas alfalfa fields) the rate would be considerably less. There are considerable regional and temporal data describing evaporation rates from pan measurements in the literature and on the web.

The rate of infiltration is the velocity at which water enters into the soil. In dry soil, water infiltrates rapidly and it is likely that in the run-off (puddle forming) events predicted by PRZM that the soil is at or near saturation. When the soil is saturated, infiltration is slower, yet reaches a steady-state rate – and these rates are fairly well predicted by soil texture (particle size and structure). The most common method to measure the infiltration rate is by a field test using a cylinder or ring infiltrometer. Measurements of water infiltration range from 30 mm/hr for sandy soil to 1 to 5 mm/hr for clay.

Use of these methods (pan evaporation and infiltration) to predict puddle duration would require estimation of puddle depth. Puddle depth, in turn, will be a function of the volume of runoff water retained on the field and surface morphology. Unless a good data set exists in the literature for such information, additional on-field observations would be required. If such information is collected, it is recommended that these data be augmented with aerial photography. Aerial photographs will not capture all puddles, but it may be the case that a correlation between large and small puddle occurrence can be determined by conducting both measurements, allowing the much larger data set available through aerial photography to be used to corroborate puddle occurrence, lifetime, and temporal

size distribution as a function of precipitation. It is possible that there may already be a large GIS data set available for this, collected by another Federal agency (e.g., NASA or the USDA) for other purposes (i.e., studies of wetlands, greenhouse gas fluxes, nitrogen loss (denitrification) in saturated agricultural soils, etc.). Infrared photography, for example, is often used to estimate soil moisture content and may correlate with puddle occurrence.

Again, an additional consideration would be to account for drinking of permanently standing water (i.e. ditches and ponds). The PRZM model already is used to account for ChemX concentrations in runoff water. This would also give consideration to exposure times beyond the half-day assumption.

Currently, the model scenarios consider precipitation only on the day after application. To provide for a more robust probabilistic exposure concentration prediction, the Panel strongly recommends using the date of application as a distributed parameter with all runoff events during the 7 day period used in the calculation of puddle occurrence. Further, regional rain gauge station information for each application scenario (i.e., Midwest corn) can be used to produce more realistic simulations of the frequency and amounts of precipitation.

For ChemX, the issue of puddle persistence may not be that critical due to the rapid rate of depuration and transformation assumed for the compound. Hence, it would not be expected that a large buildup in concentration in each bird would be predicted even with continued exposure. For other chemicals, however, this may not be the case, and continued exposure through drinking source may lead to increased body-burdens. Without considering continued exposure due to drinking water in puddles, ponds, and ditches, calculated risks will be lower than if these sources are considered.

Tillage practices will greatly influence puddle occurrence. Normal tillage into some type of mounded rows will channel and retain water into low-lying areas in the rows. If properly managed, such tillage should cause puddling whenever runoff occurs. Puddling is likely to occur before runoff initiation **and tillage such as "furrow dikes" will promote significant puddling before runoff occurs.** As a first approximation, if you have runoff, puddles should have formed. Some parameters that will control puddling are rainfall amount, rainfall duration, soil infiltration rates, plant cover, temperature, relative humidity, wind speed, and topography. Any rain event that exceeds the sum of infiltration plus evaporation will cause a puddle.

It should also be noted that there are serious data limitations that will affect the modeling of puddle formation and longevity. The meteorological data that are available provide daily rainfall total, not the time dependent intensity of rainfall

during a given day. Without better temporal resolution for rain flux onto a field, modeling puddle formation will likely be guesswork without some empirical groundtruthing as described earlier in the answer to this question. Also, to use topography as a variable, one would need to have spatial resolution to the square foot, and data of that resolutions are generally unavailable.

6. Concentrations in Drinking Water:

A. In the absence of a more rigorous model for pesticide residue in dew, EFED has used a simple two-compartment partitioning model. In the Case Study document, EFED has discussed some critical limitations to this model.

- **What model modifications can the Panel suggest for improving the estimation of pesticide residues in dew, keeping in mind the limitations of the current registration data set requirements?**
- **What any data sources regarding dew measurement or modeling should be considered to strengthen EFED's modeling needs?**

The Panel does not have a recommendation for improving the two-compartment model for estimating pesticides in dew. At present, the model is adequate for a Tier II assessment and further refinements of that equation should be reserved for more refined assessments. However, the model is overly simplistic and verification of pesticide concentrations would be advisable in situations where a significant portion of the risk is due to uptake from dew. There are two additional critical questions deserving of additional investigation: (1) how much dew is actually taken up by birds and from what source and (2) to what extent do birds in the field actually sip dew and are all birds species equally capable of sipping dew?

The Panel agreed that if ChemX is absorbed by the plant, then equilibration is unlikely to occur. However, there was disagreement as to the overall effect of kinetics on the model. This emphasizes the need for refinement if risk is deemed sufficient to proceed to a higher level PRA assessment.

One Panel member felt that a two compartment equilibrium model would considerably overestimate the concentration in the dew because of kinetic considerations that would limit transport from the carbon portion of the vegetation to the water phase. The time period allowed for the equilibrium to be established between the dew and the plant material is too short. Migration of Chem X to the surface followed by dissolution in the dew would require an extended period of time, unlike a test system where the plant material is homogenized with water and the phases are separated.

Another Panel member felt that dew formation and depletion often occur on time scales that would not allow a dislodgeable pesticide to partition into the vegetation. Achievement of equilibrium will occur on a time scale that is heavily dependent on the physical properties of the chemical in question, and may be addressed based upon whether a pesticide is considered to be systemic. Superficial/dislodgeable residues of a substance like Chem X are likely to reach equilibrium more quickly than compounds with higher K_{oc} values.

The Panel agreed that data are currently unavailable to confirm model estimates. What is needed are measurements of chemicals in dew as a means to confirm the model estimates. However, it is pointed out that residues remaining on the surface of the vegetation following spray application have not been considered in this model. If one assumed that partitioning from residue on the plant leaf surface also occurred in accordance with equilibrium partitioning, then the current approach might serve as a first approximation of the concentration that occurs in dew. In cases where the concentration of the test substance in the dew contributes significantly to potential risk, it would be advisable to verify the exposure at the next level of refinement. The closest data that the Panel could offer to help evaluate pesticide concentrations in dew are based on two lines of research: 1) studies assessing pesticide concentrations in fog (Sieber et al., 1993) and 2) studies wherein chlorinated hydrocarbon partitioning has been determined between air and corn (Wagrowski and Hites, 1998). The Panel acknowledged that the studies by Wagrowski and Hites (1998) were not presented at the SAP meeting but are being provided by the Panel as supplemental information for the Agency.

B. EFED used an instantaneous maximum approach from direct application for modeling pesticides in puddles present in a treated field on the day of application. This approach, based on experiments with buried buckets of water, does not consider partitioning kinetics with field soils.

- **What does the Panel suggest for improving this approach?**

The instantaneous maximum approach the Agency used represents a first cut conservative estimate of the amount of ChemX that might result in a puddle immediately following a spray application and, as such, is appropriate for this level of refinement. However, at higher levels, additional phenomena should be considered in model parameterization.

Partitioning between soil and water may quickly immobilize a significant fraction of the pesticide long enough to allow it to degrade or the puddle to dry up, thereby putting it outside the food chain for avian species. If the data consist of concentrations of pesticides in buckets of water, and partitioning information, then

new data are needed to generate a more realistic concentration. A possible approach would be to use the instantaneous (maximum) concentration with a two-compartment equilibrium partitioning model to estimate the concentration in the water. This would be similar to the approach that was described for calculating the concentration using partitioning to dew from vegetative matter except, in this case, the soil partition coefficient would be used with the fraction organic carbon in the soil to predict the concentration in the aqueous phase. One would expect the kinetics of this partitioning reaction to occur fairly rapidly, especially in a shallow system with 1-3 inches of water.

Field data collected under representative conditions could be used to calibrate the puddle concentrations and verify the partitioning data. For example, one could put soil in some of the buckets, making an artificial puddle, and then measure the resulting concentrations. If the partitioning rate constant is large enough, any sampling of puddles already includes the equilibrium concentration. A reason to verify the partitioning data is that at least for many compounds that are not pesticides, the reported partitioning data from laboratory experiments have been much higher (sometimes orders of magnitude higher) than what is supported by field measurements. Modification of experimental protocols in recent years has largely eliminated this apparent inconsistency.

Consideration of any time-dependent partitioning model seems to be predicated on the assumption that the lifetime of a puddle exceeds a time step in the exposure model. The models described by the Agency use a 12-hour or longer time step and use a constant concentration across this time step. A shorter time step and longer puddle life would probably be required to support the use of any time-dependent soil partitioning approach. If one moved to a time-dependent approach, one could also consider concepts such as hydrolysis or volatilization. This should allow for improved estimates of pesticides in the puddle.

It was also noted that pesticide concentrations in puddles have been reported to the Agency as part of registration packets for compounds other than ChemX. If these data include chemicals spanning a sufficiently large range of functionality and solubility, an empirically derived model could be developed. Again this solution to the modeling problem may require an agreement regarding data sharing among registrants.

C. Pesticide residues in puddles appearing on a treated field on the day after pesticide application were estimated using output from PRZM runs. The operative assumption in this approach is that water puddling on the surface of the field would be equivalent in pesticide concentration to the water potentially running off the field from a given precipitation event. EFED recognizes a number of limitations to this approach, not the least of which is

the sensitivity of the PRZM model output to application date in terms of number and magnitude of run-off event pesticide loadings following the application date.

- **What are the Panel's suggestions for adapting existing tools to this task and improving the modeling of day-after treatment puddle residues?**

The Panel agreed that the level of complexity is generally appropriate for this assessment, but the rainfall data set needs to be expanded at this level. The Agency's PRA for ChemX states (p. 23 of the Agency's background document): "For each precipitation event in the PRZM run that resulted in runoff *on the day after aerial pesticide application*, the field-wide average runoff concentration (assumed to be equivalent to average puddle concentration) was calculated. The mean value of these daily measurements (only for days where runoff was predicted to occur) was calculated and served as the mean puddle concentration on the day after application." PRZM was run with daily hydrological data over a 36-year period. Application was assumed to occur on May 6th of each year for corn, resulting in 6 runoff events over the 36-year period, and application was assumed to occur on April 1 of each year for alfalfa, with only two runoff events occurring over the 36-year period. The spatial variability was calculated from the field study puddle of variability and not from these few data points. Hence the concentration in runoff for alfalfa was calculated based on modeling only two discrete precipitation events. This is really not sufficient. As a minimum, more total precipitation events need to be considered. This can be accomplished by using rain gauge information across a region and "date of application" as either a variable or distributed variable. A data set that includes 20 regional rain gauges with 36 years of data, with a possibility of 30 different application dates leads to over 20,000 total events. Even if only two runoff events occurred per month, a database of over 1,000 runoff events would result.

The Panel also agreed that intermediate assessments should identify variables that need refinement at higher levels. Several suggestions are listed below.

(1) EPA may wish to evaluate high, medium, and low rainfall events, based on the 5th, 50th, and 95th percentile rain events over the 36-year period for the months of typical application and over the variation in regional rain gauge amounts. The 5th percentile may result in no run-off, so the 25th percentile could be used. The idea is to establish typical runoff (ponding) events during the time of application. Additionally, because the hydrological data exist and specific applications dates are modeled, it should be relatively easy to consider all precipitation events that occur over the entire 7-day period. This would allow new puddles to form beyond the 2nd day after application.

(2) Regarding the operative assumption that "water puddling on the surface of the field would be equivalent in pesticide concentration to the water potentially running off the field," the validity of this assumption depends on how accurately PRZM accounts for chemical sorption to soil and crop residues in overland flow. This may not be important for ChemX due to its low K_p, however this may not be the case for other chemicals.

(3) Currently the mean concentration in puddles contaminated by runoff is estimated from PRZM (mean of a small number of predicted runoff events over 36 years), but the variance is taken from the distribution used for simulated puddles directly treated with ChemX (which was intended to represent intra-field variability). Therefore, the resulting distribution does not represent variation between fields, or over time (i.e. between runoff events), both of which one would expect to be important. Ideally, all these sources of variation should be represented separately as the model is developed further.

(4) In the case of dew, the assumption regarding f_{oc} is very uncertain (and the value chosen is unconservative) and the K_{oc} estimates are highly variable (factor of 6). Therefore, the Panel recommends that these be varied in the model to examine their influence on risk and especially on the conclusion that dew consumption is a potentially significant contributor to risk.

7. Residues in Vegetation Food Items: In the Case Study, EFED chose to base estimates of pesticide residues on the data provided in Fletcher et al. (1994). EFED also had data on actual ChemX residues after treatment, but these data were severely limited in sample number and in a very limited number of field study areas. EFED based the decision to use Fletcher et al. on a desire to utilize a more robust, albeit non-chemical specific, data set to establish initial pesticide residues.

- **What is the Panel's opinion regarding EFED's approach, which relies on larger more generalized data sets versus focusing on limited single-chemical data sets for estimating initial field concentrations of pesticide residues**

Unless the residue data pertaining to a given pesticide are very extensive, the Panel concluded it is more appropriate to use generic data (as per Fletcher et al. 1994) than to use the chemical specific data. These generic data could be verified against empirical data at higher levels within the PRA process. Empirical data for pesticide residue variation in vegetation are available for a range of pesticides, and could be used to establish generic residue profiles. For example, Cobb et al. (2000, 2001) measured distributions of pesticide concentrations on vegetation within orchards in two geographic regions of the US.

The Panel cautions that even data contained in compendia such as Fletcher et al. may underestimate the true variance surrounding pesticide applications and, hence, residue levels in foodstuffs. The Panel agreed that the Agency should consider the large variance in residue concentration that may be manifested through variation in spray application (including applicator skill level, the specific machinery, its configuration, and the extent to which it is properly calibrated). The equipment used in supervised field trials from which the residue data were derived benefits from a degree of professional supervision and calibration that is not typical of the real world. Thus, if registrant data is used, it would be appropriate to superimpose a distribution that better reflects normal applicator to applicator variance (Hofman and Hauck [1983]; Rider and Dickey [1982]).

The Panel also cautions that estimates of pesticide distributions for invertebrates are less well established and work is needed to collate a larger database, identify patterns, partition variance, and develop robust distributions for routine use.

The Agency should also be careful not to confuse inter- and intra-field variation in drawing distributions of residue data. Currently, the data of Fletcher et al. is used for both. Each of the 20 birds in the Agency's model should be subjected to exposure levels derived from an intra-field distribution of residues. Each group of 20, on the other hand, should be exposed to mean residue levels derived from the broader inter-study distribution of the means as described by Fletcher. Those data should be reanalyzed in order to extract intra-field variance terms.

The Panel also suggested that, at higher level PRAs, the Agency consider stochastic aspects such as weather variables which contribute to a high inter-field (or inter-study) variance in residue concentrations. Such variation is usually poorly represented in the small data sets submitted by registrants. Furthermore, the variance of the residue distribution may not be independent of the mean.

One Panel member felt that the decision to use a larger, more generalized dataset, versus a limited, chemical specific data set could be questioned. The questions of concern to this Panel member were:

(1) How variable were the chemical-specific studies?
(2) How did results of the chemical-specific study compare with the generalized dataset of Fletcher et al. (1994)?
(3) Did the decision criteria include factors other than small sample size?
(4) Were there design flaws in some of the chemical-specific data?
(5) Why not use all available data?

There also was concern that the Agency was inconsistent in not using results

from the four field studies to estimate residues in food items. Instead, the Agency decided to use that same dataset to derive residue dissipation half lives values for the PRA.

8. Residue Clearance in Focal Species: EFED has included a residue retention factor in the exposure model. This factor was intended to account for carryover of a proportion of a time step's body burden to the next exposure time step, allowing for some consideration of cumulative exposure.

- **What are the Panel's thoughts on this approach?**
- **Would the Panel recommend approaches to account for uncertainty associated with extrapolating poultry metabolism data (the origin of the residue retention factor used in the Case Study) to other bird species?**
- **Please comment on any need for pharmacokinetic studies to improve this assessment approach. What are the Panel's recommendations on the scope of such studies (i.e., appropriate species, number of species, study design, and endpoints)?**

At the Tier 1 level, the Panel concluded it is reasonable to use the chicken metabolism study to predict the starting point for subsequent exposure steps, as well as to help predict peak residue values within any given step as done by the Agency. A review of the metabolism in the chicken is useful to the general interpretation of avian toxicology and, as such, should be a Tier 1 prerequisite for any pesticides where avian exposure is considered likely (currently, avian metabolism studies are available only where contaminated crop residues are to be fed to chickens.) One area where the avian metabolism study is especially useful is in the interpretation of the avian reproduction study. Birds are sufficiently different in their physiology and metabolic activity from mammals, thus mammalian rates of clearance cannot be used as proxies.

However, the Panel did not agree with all aspects and ramifications of the approach. Some of the main areas of concern are outlined below for consideration by the Agency.

Behavior Effects

As suggested by Mineau et al. 1994 (CWS Tech Rep. No. 215), the Agency could consider the LC_{50} study to estimate the toxicity of the product (Kenaga, 1993). With some classes of pesticides, it can be shown that a daily dose totaling several LD_{50}s can be ingested each day if the test subject paces its ingestion rate. Where it can be shown that, as an example 10 or 20 LD_{50} equivalents can be ingested per day before the LC_{50} is reached, it may not be reasonable to use a clearance rate obtained from the hen metabolism study. The measure of toxicity

may be easier to measure with the changes that have been proposed for the LC_{50} design. Exploratory analyses of the type presented during the public comment period for ChemX can be completed for other chemicals with different acute toxicities. This may help in establishing a threshold level below which it would be useful to account for the additive nature of exposure from one time step to the next.

Inter-specific variation

As far as assuming that the chicken is representative of all bird species, there is clearly a large uncertainty. It may be possible to obtain a variance estimate for interspecies differences by mining the literature. Unfortunately, most of the data are likely to be for persistent organochlorines and metals and may be of limited applicability for other chemical classes. Metabolism may be tied less to phylogeny than to diet for both inter- and intra- species. One Panel member also expressed reservations about the fact that the chicken has a well developed crop ensuring that ingested food is processed gradually over time. This is not the situation with small passerines that don't have a well-developed storage organ.

In summary, the Panel acknowledged that different species will vary in their clearance rates but, at this juncture, it is probably best to compare model outputs with and without (chicken-based) metabolism to determine whether inter-species differences in metabolism are likely to make a difference to the final model output. It is also clear that even if rates of elimination could be obtained for species of interest, these would likely vary with season, diet, etc. The Panel suggests that information obtained from *in vitro* liver cultures (measurement of phase one and phase two enzymes) as well as the mammalian metabolism data, may help in estimating inter-species metabolic differences. The Panel recognizes that this is a research approach that cannot be pursued in the short-term by the Agency or the scientific community.

Assumption of first-order kinetics

Several Panel members concluded that pesticides are unlikely to follow first-order clearance rates and recommended that a more plausible model be developed. A general two or three compartment model (e.g., lipid pool - plasma - target organ) should be developed, drawing on the proper expertise in this area. Any model should strive to obtain a better measure of internal dose. It is probable that the chicken metabolism study will have to be expanded (i.e., new compartments would need to be measured) if such a model is created. The Panel recognized that a move away from first order kinetics would make some of the exposure parameters (e.g. the intake rate) much more important than they are at present.

As presented during the public comments at the meeting, allowing body

burdens to accumulate from time step to when they reach a critical threshold is a new way of characterizing exposure and needs to be carefully considered. This new method of comparing exposure to effect may lead to a very conservative (over-protective) assessment of the risk if body clearance is slower than clearance from active sites of toxicity.

Validation of the paradigm

For chemicals with a well defined mode of toxic action (e.g., cholinesterase inhibition), it may be more logical to look at the active site of toxicity (e.g., recovery from brain ChE inhibition) rather than clearance from the whole body. This would be especially applicable where pesticides are metabolized to even more toxic moieties (e.g. the organophosphorus insecticide acephate). The corollary is that a reliable biomarker of a pesticide's toxic mode of action should be available. The Panel is aware that such biomarkers may be less and less available as we move away from ChE inhibitors and embrace new chemistry. Thus, the Panel proposed that the Agency consider developing data requirements for identification of reliable biomarkers. The ideal biomarker would be one that is closely tied to the compound's toxic mode of action.

9. Selection of Exposure Time Steps: The Case Study divides each exposure day into two time steps. There have been questions regarding the effect of exposure time step and feeding duration upon the outcome of the model. To investigate this, EFED used a mass balance equation with two compartments, to determine the effect of rate of food consumption combined with clearance rate, on total body burden (ingested + residual) of pesticide. The results of this exercise led EFED to conclude that duration of the exposure window within a time step had minimal effect on maximum instantaneous body burden of ChemX.

- **Given the lack of detailed pharmacokinetic data for ChemX in birds, what other analyses of this issue could be made?**

With regard to avian feeding behavior, the use of two time steps per day is overly simplistic. Feeding by most birds can occur throughout the day, but there would be two peaks in food consumption, one peak occurring shortly after sunrise (after an overnight fast) and a second peak before sunset. There is a lull in feeding in the early afternoon due, in part, by high midday temperatures.

A time-step model as presented by the Agency is reasonable as a Tier 2 approximation and should produce credible output provided:1) pesticide intake does not affect the continued rate of intake (no avoidance - either conditioned aversion or post-ingestional feeding incapacity) or 2) avoidance occurs too late

relative to an intake commensurate with lethality. There is growing evidence that the rate of pesticide intake is key to an individual's probability of death in the case of highly toxic pesticides (Hart et al., 1999).

Clearly, the effective feeding rate within a time step does matter in real life even if this is not currently captured by the model. The critical variable with respect to ChemX and other highly toxic pesticides may be the size of the meal rather than the maximal body burden attained over the course of a time step. (Note that, under this approach, meal sizes could be based on gut content or on intake-to-satiation data from optimal foraging models). However, the model need not represent actual feeding processes; rather it needs only to provide a reasonable assessment of the risk.

There are other ways in which the model does not emulate the feeding behavior of real birds. The actual feeding time within a time step was modified by the Agency in a sensitivity analysis but independently of the biological reality of foraging birds. In fact, it would be unusual for a food intake scenario to be biologically plausible for all time intervals between 1 min. and 12 hours as indicated in table 26. It is likely there will be constraints on both ends of these extremes - gut volume and food handling time at one extreme and costly foraging strategy at the other. These topics could be addressed in higher level PRAs.

The Panel asks that the Agency reevaluate the calculations that purport to show that the exposure window within a time step has a minimal effect on maximum body burden. The results as shown are counter-intuitive and the calculations appear to be in error. A shorter effective exposure window, when placed at the beginning of the time step, should result in more efficient clearance of the pesticide by the end of the time step.

The Panel suggests that Agency consult with its colleagues at it's National Health and Environmental Effects Research Laboratory for an assessment of metabolism data. Other experts in avian biochemistry may have insights regarding relative metabolic capabilities of chickens relative to other avian species (Grau and Wilson, 1964; Wilson et al, 1969; Entrikin et al., 1977; Entrikin et al., 1988; Wilson, 1990). Finally, developmental profiles of metabolic enzymes have been developed for some terrestrial species and should be evaluated to better understand susceptibility of nestlings.

10. Selection of Acute Toxicity Standard: In the Case Study, EFED used the acute single oral dose studies (LD_{50}) as the basis for characterizing effects in preference to the acute dietary exposure studies. Both of these study designs have limitations for estimating the risk to wild avian species exposed to pesticides in the environment. Both studies have a fixed exposure period, not

allowing for the differences in response of individuals to different duration of exposure. For the acute oral study, the dose administered in a single dose all at one time does not mimic wild birds' exposure. Also, for exposure through different environmental matrices, it does not account for the effect of the matrices on the absorption rate of the chemical into the animal.

This latter criticism also applies to the dietary test for other food matrices consumed in the wild. For the dietary test, the endpoint is reported as the concentration mixed with food that produces a response rather than as the dose ingested. There are a number of study limitations that render conversion of dietary concentration to dose problematic (e.g., food spillage, and quantifying food intake for individuals). The interpretation of this test is also confounded because the response of birds is not only a function of the intrinsic toxicity of the pesticide, but also the willingness of the birds to consume treated food.

More importantly, there is evidence for some compounds that the laboratory derived LC_{50} values are poor predictors of effects in the field.

- **What are the Panel's thoughts and recommendations regarding alternative approaches using LC_{50} data that allow estimation of dose-response relationships (critical to prediction of magnitude of effects)?**

The Panel believes that the Agency was correct in choosing to use the LD_{50} over the LC_{50} in the case of ChemX and wherever there was any demonstrated food aversion in the LC_{50} test. Mineau et al. (1994) showed that, as a result of food aversion and other problems with the test, LC_{50} endpoints were often inconsistent from test to test, species sensitivity 'flip-flopped' to the extent that no one species was a good predictor of birds in general neither did the LC_{50} endpoint appear to have much field relevance. Typically, birds in the laboratory are able to pace their dietary intake to avoid toxicosis while much lower levels cause mortality in the field.

LC_{50} tests need not be discounted if food avoidance is not an issue. The usefulness of the test may increase especially if its design is improved to more easily measure food consumption, control bird age, size and condition, and measure the caloric value of the feed as proposed by an OECD expert group. Once the concentration of pesticide in the feed is corrected on the basis of caloric content (e.g., mg pesticide/kcal of diet) and the test birds are treated in a manner more consistent with their wild counterparts (e.g., slight food stress, shorter feeding intervals and/or perceived competition for the food), it would be easier to extrapolate from the laboratory to the field (Hart et al., 1999). Of course, this assumes (as the Agency assumes currently) that the caloric content of food is the

'proper' currency perceived by foraging birds. In actual fact, birds may also adjust their feeding according to protein or even mineral content. A further difficulty with any future use of the LC_{50} in probabilistic assessment is that there would not be a good basis on which to base interspecific sensitivity differences. Indeed, the maximum number of LC_{50s} available for any given pesticide is likely to be 3 or 4 at the most. The LD_{50} interspecies variance would probably need to be superimposed onto the LC_{50}.

Of course, the LC_{50} would be more relevant if demonstrated feed aversions could be easily extrapolated to the field, which is not currently the case. All evidence to date indicates that the current LC_{50} test as well as simple choice tests are much more likely to trigger avoidance than is the case in the wild. It has even been shown that the exact shape of the food dish could alter the ease with which conditioned food aversion can be established in the laboratory (Fryday et al, 1998). ChemX is a very acutely toxic insecticide. This is yet one more reason why the LD_{50} test is the most reasonable choice. A simple calculation (see table 26 of the Agency's background document) indicates that very short feeding intervals at reasonable ingestion rates can cause mortality.

As suggested by the ECOFRAM committee, the Agency is correct in paying attention to scaling of the LD_{50}s to account for differences in bird size following the work of Mineau et al. (1996). In general, for the majority of pesticides, not scaling for size may seriously under-protect small species. ChemX appears to differ from the majority of pesticides in that it scales to less than unity. The scaling factor specific to ChemX leads to the assumption that larger bodied birds are more sensitive than smaller bodied birds. There were concerns expressed by the Panel about the application of the scaling factor in the case of ChemX. Scaling to body weight does not remove any species sensitivity relations that arise as a result of phylogeny. Waterfowl (some of the larger species) appear to be particularly sensitive to ChemX. The ring-necked pheasant, the largest species tested one of the least tested and least sensitive gallinaceous birds in general, appeared less sensitive. This results in considerable scatter in the data points describing the relationship between LD_{50} and body weight, suggesting that the R^2 for the scaling factor is low. When there is obvious scatter in the size-sensitivity relationship, especially where there are numerous data points, the Agency should explore the various options that are open in order to best characterize interspecific sensitivity differences.

The Panel did not agree with the Agency that data points should be thrown out because ranges are given. When compiling data for Mineau et al. (2001), the authors were advised that ranges given in various publications by Schaefer et al. (1983) actually referred to repeat up and down testing rather than to the uncertainty within any given test. In either case, the geometric mean could be used as an

adequate approximation of the lethal dose. Data that provide information on interspecies sensitivity should be used to their fullest extent

It was suggested by one Panel member that the raw data from the test results should be obtained and presented where possible. This would allow for the exploration of alternate models for toxicity determination. It was also suggested that the bounds of the toxicity distribution might be adjusted to reflect a finite species pool (to represent potentially exposed species rather than all bird species).

The Panel noted that LD_{50} and LC_{50} exposure studies have limitations for estimating the risk to wild avian species exposed to pesticides in the environment. Both studies have a fixed exposure period, not allowing for the differences in response of individuals to different durations of exposure. For the acute oral study, the dose administered in a single dose all at one time does not mimic wild birds' exposure. Also, for exposure through different environmental matrices, it does not account for the effect of the matrices on the absorption rate of the chemical into the animal. This latter criticism also applies to the dietary test for other food matrices consumed in the wild. For the dietary test, the endpoint is reported as the concentration mixed with food that produces a response rather than as the dose ingested. There are a number of study limitations that render conversion of dietary concentration to dose problematic (e.g., food spillage and quantifying food intake for individuals). The interpretation of this test may be confounded because the response of birds is not only a function of the intrinsic toxicity of the pesticide but also the willingness of the birds to consume treated food.

One Panelist directed the Agency to an avian workshop where the studies that should be required as a basic minimum data set were discussed. The workshop participants agreed that, subject to certain conditions and exceptions, a basic minimum of a single acute toxicity study and a single reproductive toxicity study might be sufficient for the initial assessment of pesticides. The workshop proposed an assessment framework in which additional studies, potentially including dietary studies or additional acute and reproductive studies, would be conducted only when a need for them is identified by an initial risk assessment conducted with the two basic studies (Hart et al., in press).

11. Defining the Distribution of Species Sensitivity: EFED has selected a combination of methods to establish the distribution of species sensitivity to ChemX, from which representative points, low, medium, and high sensitivity are used to characterize the uncertainty regarding response of any particular species to ChemX exposure. The first step in the method involves normalizing all available toxicity data to a constant body weight using weighting factors established for specific and generic chemicals. The mean and standard deviation of these normalized values is taken. Then the method of Aldenberg

and Slob (1993) is used to estimate the 5th and 95th percentiles of a log-logistic distribution with mean and standard deviation as defined above. Finally these point estimates of sensitivity are readjusted to a focal species body weight.

- **What are the Panel's views on this method to estimate the distribution of sensitivity of focal species when species-specific toxicity data are not available?**
- **How would the Panel estimate the confidence interval surrounding the above estimates of the 5th and 95th percentiles?**

The Panel commends the Agency for attempting to use available data to justify the choices of values for the low, medium, and high sensitivity scenarios and for incorporating a sensitivity distribution approach. This is a crucial component of the PRA because of the large apparent variability between species and the high dependence of the calculated risk distribution on the choice of species sensitivity. There are numerous suggestions for possible inclusion in the Agency's approach:

1) The conceptual issues are similar to those for the frequency of birds in treated fields. To what extent do the 15 species for which there are LD_{50} data provide information about the unknown LD_{50}'s for bird species on the treated field. There are two extremes:

- The 13 species are a random sample from the population of species of concern.
- Each of the 13 species is a unique case because the biology of each species is unique.

2) Selecting the minimum and maximum LD_{50}'s observed among the 13 (or 15) species gives similar LD_{50}'s for low and high scenarios. Hence, the details of how the values are determined are less important than communication of the method and acceptance that the method is reasonable.

The Agency's approach is a compromise: species differ in body weight, which influences the species LD_{50}. As the Agency commented, the proportion of variance explained by the scaling with body weight is not very high. Thus, there is substantial uncertainty in the scaled values. This uncertainty probably should be incorporated into the confidence limits on the estimated LD_{50}s for focal species. However, the similarity in the bounds for different species suggests that the influence of body weight is minor if the current form of body weight adjustment is correct. The body weight adjusted values are treated as a sample from the unknown distribution of LD_{50}'s. The observed LD_{50}'s from 15 species are consistent with both log logistic and log normal distributions. They are probably consistent with many other distributions because the sample size is small.

Biological knowledge is useful to help choose a distribution.

The Agency assumes that test species are representative of wild species. This assumption is probably biased, but the extent and effect of bias is uncertain. Given the high influence of sensitivity on the risk estimates, this uncertainty deserves further discussion and investigation (e.g., looking at the underlying data to see whether estimates based on a subset of species are representative of the rest and incorporating any estimated bias into the model as uncertainty to judge its effect on the risk estimate). One Panel member differed, stating that most of the birds species tested for Chem X and other pesticides with large data sets are wild birds brought in to the laboratory for a short acclimation period.

3) The estimated bounds are likely to be too wide (i.e., too small for the 5th percentile and too large for the 95th percentile), because the bounds are calculated from estimated quantities (LD_{50}s) not the unknown distribution of 'true' species values. The uncertainty in the LD_{50} for a species is included in the species-species variability of LD_{50}s. The confidence intervals in Table 14 illustrate the sometimes large uncertainty in a single estimate of LD_{50}. Table 14 also illustrates the variability between age groups. This variability appears to be much smaller than the variability among species. Hence, the current approach of averaging over age groups to compute a species mean value is appropriate. A variance components analysis to separate estimation error (uncertainty) from species-species variance could be used.

4) Aldenberg and Slob's (A&S) method treats the LD_{50} values as point values, but in fact they often have fairly wide confidence limits. Aldenberg and Jaworska (2000a, 2000b) have developed refined methods for estimating confidence limits for percentiles of the species sensitivity distribution (SSD), taking account of the uncertainty in the test results on which it is based as well as accounting for uncertainty due to the number of species tested. The Panel suggests that the Agency consider trying these approaches at least to see how much difference they make.

5) The Panel suggests that data be used from all the studies. The distinction between values for different ages (used as geometric mean) and values for different studies (omitted from analysis) seems arbitrary. The Panel suggests using the geometric mean of the different studies.

6) The A&S extrapolation factors are confidence bounds (one-sided confidence intervals), so they incorporate the uncertainty in the estimated mean and standard deviation. If the population mean and variance were known exactly, the 5'th percentile of the log-logistic would be given by mean – 1.62 * sd. The value in A & S's table for the 5th percentile are the 50% and 95% lower confidence bounds

for the 5th percentile. These values incorporate the uncertainty in the estimation of the mean and variance. The extrapolation factor used by the Agency, 1.71, is the coefficient for the 50% lower confidence bound on 5'th percentile for a sample size of 13. This bound is relatively close to the coefficient for known mean and variance (1.62).

In the statistical literature, the quantities being computed are called tolerance intervals. For normal or lognormal distributions, the quantities needed can be computed without simulation. Many of the quantities are tabulated. Extensive tabulations are in Odeh and Owen (1980). Less extensive tables are in Gilbert (1987) and Hahn and Meeker (1991). Both Gilbert (1987) and Hahn/Meeker (1987) describe the use and derivation of these intervals. The theory is rigorously given in the introduction to the Odeh and Owen (1980) tables.

The current usage of the extrapolation factors is not statistically consistent, but the practical effect of this is limited when using the 50% confidence bound. To illustrate this, the high, medium, and low sensitivity values are described verbally. The verbal descriptions are expressed in terms of the LD_{50} values for a randomly chosen species. The current values are described below:

Low: The 50% lower confidence bound to the 5th percentile (50% confident that 95% of the species [i.e. 100-5%] have LD_{50}s larger than this value).

Medium: The 50% lower confidence bound to the 50th percentile (50% confident that 50% of the species have LD_{50}'s larger than this value). Because of the symmetry of the logistic distribution, this is also the 50% upper bound (50% confident that 50% of the species have LD_{50}s smaller than this value). This value is also the 50th percentile if the mean and variance are known precisely.

High: The 50% upper confidence bound to the 95% percentile. You are 50% confident that 95% of the species have LD_{50}'s less than this value.

95% confidence bounds (1-sided intervals) can also be employed. A decision will need to be made whether an upper 95% bound or lower 95% bound for the high scenario is needed. These bounds would be as follows:

Low: value currently calculated.

Medium: 95% confident that 50% of the species have LD_{50}s larger than this value.

High: (lower bound) 95% confident that 5% of the species have LD_{50}'s larger than this value.

High: (upper bound) 95% confident that 95% of the species have LD_{50}'s smaller than this value.

The coefficients can be estimated using the same simulation approach as A&S (1993). For 50,000 replicates of samples of n=13 species, they are:

95% lower bound for the 5th percentile: k = -2.799 (similar to A&S's value)
95% lower bound for the 50th percentile: k = -0.50 (similar to A&S's value)
95% lower bound for the 95th percentile: k = 1.007
95% upper bound for the 95th percentile: k = 2.799

7) The literature on tolerance intervals also includes non-parametric tolerance intervals. These estimate the same quantities (e.g,. a 95% lower bound on the 5th percentile of a distribution) without assuming a specific distribution. They do require much more data. A lower 95% bound to the 50th percentile can be estimated non-parametrically from five or more observations. The lower 95% bound to the 5% percentile cannot be estimated non-parametrically unless there are 59 or more observations. The non-parametric tolerance intervals are analogous to a non-parametric bootstrap estimate of the confidence bounds. The non-parametric bootstrap cannot estimate the desired quantiles, because of the small sample size (number of species with estimated LD_{50}'s).

8) The tolerance interval literature suggests another approach. The currently computed values describe inferences about an infinite population of species. The Panel's opinion is that the species pool is likely to contain a relatively small number of species, e.g., 10 or 20. Assume for the following that the species pool contains 10 species, none of which have a measured LD_{50}. Each of these 10 species has an LD_{50} that is randomly chosen from the uncertain distribution of LD_{50}'s. It is possible to calculate X with the properties that:

Low: 95% confident that all 10 out of 10 randomly chosen LD_{50}'s are larger than X.

Medium: 95% confident that 5 out of 10 LD_{50}'s are larger than this value.

High: 95% confident that 0 out 10 LD_{50}'s are larger than this value.

The extrapolation factors are quite different for the upper and lower bounds to the 95% quantile.

The extrapolation coefficients for these intervals depend on the size of the species pool. They are larger for a 10 species pool than for a 6 species pool. These values are tabulated for normal distributions (e.g., Odeh and Owen, 1980).

9) Some of the LD_{50}s result from approximate lethal dose (ALD) tests in which the choice of dose levels constrains the values taken by the reported endpoint. This probably explains why the value 1.33 appears in the results for 4 of the 15 species. The influence of this issue on SSDs needs examining.

10) There are a number of other questions concerning the construction and use of species sensitivity distributions. These issues were reviewed in an avian effects assessment workshop report (Hart et al., in press). These include:

What distributional form best describes variation in species sensitivity? Should there be an assumption of a particular distribution or use of a bootstrapping method?

Are effects correlated between different species?

Are avian and mammalian effects correlated?

To what extent is between species variation merely a reflection of between test variation for a single species?

Is it appropriate to use a pooled estimate of variance from a number of different compounds?

Is inter species variance related to size of test?

Can we use historic data to refine inferences about new compounds?

What is the appropriate balance: number of species versus size of test?

How should we deal with censored data?

As discussed in earlier questions, once the Agency establishes distributions of this type they are going to be used repeatedly. Thus, there is a need to ensure proper development. In the case of SSDs, this means addressing the questions above. The SETAC avian workshop recommended that a collaborative effort be established for this purpose. Given the high influence of SSD on the risk estimates, the Panel suggests that such an effort should be a high research priority for the Agency.

(11) The approach taken by the Agency to estimate the distribution of sensitivity of focal species when toxicity data are not available should be considered in light of what is actually known about the species sensitivity. First, the calculation of the 5th

percentile in a SSD does not necessarily correspond to a given species. It is an estimate of the concentration where 95% of the species would be expected to have oral LD_{50} values greater than the estimated value. Second, one should not assume that a series of focal species for which there are no data are equally sensitive to the 5th, 50th, or 95th percentile. The concept of using the SSD is that it provides a description of the range of sensitivity that exists for birds to Chem X. The 95th percentile, by convention (accepted standard practice) is a concentration that is thought to be sufficiently low to be protective of most species. To assume a given number of focal species all having sensitivities equal to the 95th or some other percentile without accompanying data, this would not fit distributional theory (i.e., it is nearly impossible for them to all be equally sensitive).

However, the Agency can conduct a probabilistic risk assessment for a generic species that has sensitivity equivalent to the 95th percentile. To correct for differences in species weight, a distribution of weights could be used to accompany the species at the 95th percentile.

(12) One Panel member believed that the Agency should obtain raw data for LD_{50}'s and use consistent estimates of LD_{50} and confidence intervals.

12. Other Factors Affecting Sensitivity: By relying on laboratory LD_{50} data, we have not accounted for a variety of physiological and environmental factors that may modify sensitivity (e.g., age, nutritional status, temperature, etc.) . EFED has concerns that not accounting for a number of these variables may lead to an underestimation of risk.

What guidance can the Panel offer that would allow, in the absence of chemical specific data, for a consideration of these physiological and environmental factors?

Physiology

The Panel discussed a variety of physiological data to some extent in answer to question 8. The uptake rate across the gut, metabolism and excretion are primary parameters to be addressed. Temperature also affects nutritional and water requirements for animals. This may be appropriate at higher levels of refinement.

There is also an age dependent toxicity that should be evaluated. It is well documented that in young birds, many metabolic and protective enzyme systems are not fully developed. Impacts on young organisms are most likely to cause population level effects (population modeling). Thus, evaluating dose responses for young birds compared to older birds would be appropriate at this level of refinement.

Behavior

One Panel member commented that avoidance and regurgitation could be important in reducing risks. Both occurrences depend greatly on time scale. Factors affecting avoidance during short bursts of rapid feeding are very different from those affecting avoidance over periods of hours or days. Both could reduce the risk very significantly. However, simplistic analysis of these factors could severely underestimate risk – e.g., birds may completely avoid treated food in a simple choice test and yet the treated food can cause substantial mortalities in the field. These factors also have implications for the structure of the model. The Agency will need a much shorter timescale for feeding in order to model avoidance in the rapid feeding situation.

Environmental Factors

Degradation rates may be dependent on environmentally relevant temperature fluctuations (280K-315K or 7C-42C). Temperature dependence takes an exponential form. There is an energy of activation term that can be predicted through molecular modeling for the critical functional groups within a class of molecule. For long lived compounds, some fraction may overwinter as degradation rates in freezing or frozen substances are negligible. This is a consideration that may be appropriate in higher tier assessments.

Toxicant availability is greatly controlled by partitioning onto soil or onto organic matter of food. This affects transport across the intestine (as noted above). Given the log K_{oc} for ChemX, this may not be critical, but for the general risk assessment process, there may be a large variation in chemical uptake that is dependent on the composition of material in the stomach of the organism consuming the pesticide.

Formulation of the compound may facilitate uptake or activate metabolic pathways that convert the compounds in question to more toxic compounds. Uptake evaluations may be more easily performed than evaluation of important activation processes. Evaluating uptake rates or altered metabolic activity will require additional data collection. Thus, incorporating this parameter would be appropriate only in higher level PRAs.

OTHER COMMENTS

The Panel also provided additional comments in its review of the Agency background and other issues for consideration. The comments are listed below.

The Panel found the use of the term ‘average risk’ confusing. Would a term

such as ‘probability of mortality’ be more precise?

The Agency’s document, 'Guiding Principles for Monte Carlo Analysis,' which was provided to the Panel, contains many excellent suggestions which are relevant to both the conduct and reporting of probabilistic assessments. The Panel urges the Agency to consider adopting more of these suggestions in future phases of its work.

The Agency compares body burden to the LD_{50} when integrating exposure and effects to estimate risk. But the LD_{50} is expressed in terms of external dose. If the LD_{50} were expressed in terms of internal dose it would be lower, so the risk estimate would be higher. The Panel was uncertain how much difference this would make (it is likely to depend on the pharmacokinetics of the compound) but suggests it needs investigating if the Agency is going to continue with the approach of using body burdens.

It would be very helpful if the document included a table summarizing the parameters included in the model, showing for which of them variability and incertitude were modeled and at which level of the model they are varied (between time steps, individuals, cohorts/sites, species).

There is almost no discussion of dependencies among variables, even though there are some parameters in the model that are clearly likely to be inter-dependent (e.g. residues on different food items).

Several Panel members expressed concern that the Agency was using the Nagy equations to vary field metabolic rate (FMR) within species, as a function of body weight. Nagy’s equations describe between species variation and are poor predictors of intra-species variation.

The Agency produced three main types of output: tabulated mortalities, binomial curves for individual species, and a composite curve (Figure 3 in the Agency’s background document). It was questioned whether these outputs have a sufficiently meaningful interpretation in real-world terms (e.g., are they intended to refer to cohorts of 20 birds on 1,000 realizations of one site, or are they intended to represent cohorts of 20 birds on a sample of 1,000 different sites?). These issues have significant consequences for the structure of the model. It was suggested that risk managers should be consulted to determine what types of outputs would best meet their needs.

The third type of output mentioned above (Figure 3 in the Agency’s background document) is an interesting attempt to combine the results of modeling a variety of different species and scenarios, but this output is rather difficult to interpret. Among other things, it confounds variability (e.g., between species) and

uncertainty (e.g., between scenarios), whereas the EPA guidance on Monte Carlo suggests these should be kept separate.

Finally, one Panel member expressed reservations regarding the utility of testing the risk assessment paradigm with ChemX. The acute toxicity and relatively short life of ChemX were primary reasons for concern. The factors for continued exposure over time did not play much of a role in the model. Many birds died quickly and residual effects were minimal. The model would have been given a better test with a more persistent compound, although the present model did show that birds would die with the product studied.

REFERENCES

Aldenberg, T. and Jaworska J. 2000a. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. Ecotoxicology and Environmental Safety 46: 1-18.

Aldenberg, A., Jaworska J., and Snell T. 2000b. Estimation of HC5 taking into account uncertainties of individual dose response curves and species sensitivity distribution. Presentation given at SETAC 21st Annual Meeting, Nashville, 12-16 November 2000.

Best, L. B. 1986. Conservation tillage: Ecological traps for nesting birds? Wildlife Society Bulletin 14:308-317.

Best, L. B. 2001. Temporal patterns of bird abundance in cornfield edges during the breeding season. American Midland Naturalist: in press.

Best, L. B., R. C. Whitmore, and G. M. Booth. 1990. Use of cornfields by birds during the breeding season: the importance of edge habitat. American Midland Naturalist 123:84-99.

Cobb, G.P., L.W. Brewer, E.H.H.. Hol, and C.M. Bens. 2001. Diazinon dissipation from vegetation in apple orchards from Pennsylvania and Washington. in ACS Symposium Series: Field Dissipation of Agrochemicals, E. Authur ed., ACS Press: Washington, DC (in press).

Cobb, G.P., Bens C.M., Kendall R.J., Mellott R.J, and Brewer L.W. 2000. Diazinon dissipation from vegetation and occurrence in earthworms and avian GI-tracts Collected from apple orchards following D-Z-N® 50W application. Environ. Tox. Chem. 19(5):1360-1367.

Crocker D.R., Prosser P., Bone P., Irving P., Brookes K. 2001. Project PN0915:

Improving estimates of wildlife exposure to pesticides in arable crops. Milestone report 03/03. Radio-tracking progress report. Central Science Laboratory, York, UK.

Entrikin, R.K., Abressch, R.T., Bradford, D.P., Larson, D.B., Longley. K.J., and Wilson, B.W. 1988. Glucocorticoids in muscular distrophy: beneficial effects of dexamethazone on avian myopathy. FASEB J, 2:2722-2725.

Entrikin, R.K, Swanson, K.L., Weidoff, P.M., Patterson, G.T. , and Wilson, B.W. 1977. Avian muscular distrophy: functional and biochemical improvement with diphenhydratonin. Science. 195:873-875.

Frawley, B. ., and L. B. Best. 1991. Effects of mowing on breeding bird abundance and species composition in alfalfa fields. Wildlife Society Bulletin 19:135-142.

Fryday, S.L., S.A. Chandler-Morris & A.D.M. Hart. 1998. Effect of presentation method on the avoidance of fonofos-treated seed by captive birds. Bulletin of Environmental Contamination and Toxicology 61: 448-454.

Gilbert, R. O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Grau, C.R, and Wilson B.W. 1964. Avian oogenesis and yolk deposition. Experimentia. 20:26.

Hart A., Balluff D., Barfknecht R., Chapman P., Hawkes A., Joermann G., Leopold A., and Luttik R. (In press). Avian Effects Assessment: A Framework for Contaminants Studies. SETAC Press.

Hart, A., Fryday, S., McKay, H., Pascual, J. & Prosser, P. 1999. Understanding risks to birds from pesticide-treated seeds. In: Adams, N. & Slotow, R. (Eds), Proc. 22 Int. Ornithol. Congr.. Durban: pp. 1070-1087. Birdlife South Africa, Johannesburg.

Hahn, G. J. and Meeker, W. Q. 1991. Statistical intervals, a guide for practitioners. Wiley, New York.

Hofman, V. and Hauck, D. 1983. Sprayer check in North Dakota. Amer. Soc. of Agricultural Engineers. Paper No. NCR-83-404.

Hommes, V. M., Büchs, W., Joermann, G., and Siebers, J. 1990. Vogelgefaehrdung durch Pflanzenschutzmittelrueckstaende in Blattpfuetzen auf

Gemuesekohl [Transl.: Poisoning risk to birds from residues of pesticides in leaf puddles of cole crops] Nachrichtenbl. Deut. Pflanzenschutzd. 42:113-117.

Kendall, R.J., Brewer L.W., Hitchcock, R.R., and Mayer, J.R. 1992. American wigeon mortality associated with turf application of diazinon AG500. J Wildl Dis. 28: 263-7.

Kenaga. 1993. Environ. Qual. Saf. 2:166-181.

Mineau, P., A. Baril, B.T. Collins , J. Duffe, G. Joerman, and R. Luttik. 2001. Reference values for comparing the acute toxicity of pesticides to birds. Reviews of Environmental Contamination and Toxicology 170:13-74.

Mineau, P., B.T. Collins and A. Baril. 1996. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Regulatory Toxicology and Pharmacology 24:24-29.

Mineau et al. 1994. Canadian Wildlife Service Tech. Rep. No. 215.

Mortensen SR, Chanda SM, Hooper MJ, and Padilla, S. 1996. Maturational differences in chlorpyrifos-oxonase activity may contribute to age-related sensitivity to chlorpyrifos. J Biochem Toxicol. 11: 279-87.

Odeh, R.E. and Owen, D.B. 1980. Tables for Normal Tolerance Limits, Sampling Plans, and Screening. Marcel Dekker. New York.

Patterson, M. P., and L. B. Best. 1996. Bird abundance and nesting success in Iowa CRP fields: the importance of vegetation structure and composition. American Midland Naturalist 135:153-167.

Rider, A.R., Dickey, E.C. 1982. Field evaluation of calibration accuracy for pesticide application equipment. Transactions of the ASAE 1982;25:258-260.

Seiber, J.N., Wilson, B.W., and McChesney, M.M. 1993. Air and fog deposition residues of four organophosphate insecticides used on dormant orchards in the San Joaquin Valley, California. Environ. Sci. Technol. 1993. 27: 2236- 43.

Wagrowski and Hites. 1998. Partitioning of polychlorinated dibenzo-p-dioxins and dibenzofurans between the atmosphere and corn. Environ Sci Technol. 32:2389-2393

Wilson, BW. 1990. Developmental and maturational aspects of inherited avian myopathies. Proc Soc Exp Biol Med. (2): 87-96

Wilson, B.W, Mettler, M.A. Asmundson RV. 1969. Acetylcholinesterases and non-specific esterases in developing avian tissues: distribution and molecular weights of esterases in normal and dystrophic embryos and chicks. J Exp Zool. 172 (1): 49-58.

Probabilistic Models and Methodologies: Advancing the Ecological Risk Assessment Process in the EPA, Office of Pesticide Programs

A Probabilistic Model and Process to Assess Risks to Aquatic Organisms

CHARGE

1. Exposure Model Input Parameter Variability. In addressing the regional effects to the farm pond, we have used the exposure model matrix by looking at the combination of 3 pHs, 3 field-to-pond size ratios, and 2 soil aerobic metabolism rates, without changing the meteorological data. We are currently pursuing development of an exposure model to include the following parameters as variable inputs into PRZM/EXAMS: field/pond size, Kd, soil aerobic metabolism, application date, pond depth, and pH. What other parameters should be considered as variable distributions? Would the Panel please list other recommendations or suggestions for refining this approach, considering our purpose of PRA?

2. **Exposure Distribution Profile Selection.** The exposure component of the aquatic risk assessment model uses 36 year rainfall data to generate 36 annual maxima for exposure concentrations. Two approaches were employed in establishing an exposure distribution profile: a theoretical fitted distribution using Monte Carlo analysis and an empirical distribution using a bootstrap method. Both methods performed similarly except in the tails of the distributions. The empirical distribution is preferred due to its objectivity and speed of calculation. What should the criteria be for choosing between theoretical and empirical methods?

3. **Interspecies Variability.** In developing a sensitivity curve for freshwater fish with the available data, species' data were combined into their respective families. This was done because all families except salmonids had a single representative species, whereas, salmonids had four representatives. The aim was not to skew the sensitivity data by the over-representation with salmonids. The geometric mean of the multiple species and/or multiple tests with the same species was used in establishing points along this curve. What does the Panel think of this approach? Please provide alternative recommendations, if any, for dealing with limited data sets.

4. **Effects Input Distribution.** The extrapolation of fish sensitivities used a

lognormal distribution of toxicity (LC_{50s}) and a normal distribution of the dose-response slopes. What does the Panel think of this approach and which other approaches could have been used?

5. **Extrapolation with Limited Data.** Since only one acceptable toxicity test was available for an aquatic invertebrate, an extrapolation using toxicity profiles of other compounds in the same pesticide family to determine the average sensitivity of the tested species and extrapolate a species sensitivity distribution was employed. What does the Panel think of this approach? What are the Panel members opinions on alternative approaches that may be used for establishing a sensitivity profile across diverse invertebrate taxa when one or a very few tests are available?

6. **Taxa Aggregation.** Freshwater taxa were separated from marine taxa in this case study. Since the marine data sets were limited to a single test species, toxicity profiles for that species were used in the assessment and no sensitivity distribution across taxa performed. Data from other related pesticides for marine species were also not available as in the case for freshwater invertebrate taxa. What is the Panel's opinion of separating toxicity data from marine organisms from data on freshwater organisms for purposes of establishing sensitivity profiles? Would the Panel please provide an alternate recommendation, if it has one?

7. **Chronic Assessment.** Available chronic data were limited and, therefore, fewer scenarios were considered sufficient to cover the range of outcomes. An exceedence probability approach was taken to evaluate the potential of chronic effects. What alternative approaches to evaluating chronic effects could have been taken?

8. **Estimation of Species Sensitivity.** In this case study, probabilistic assessments were performed for specific species (e.g., bluegill sunfish and rainbow trout) and for extrapolated species (e.g., 5^{th} percentile sensitive and 50^{th} percentile sensitive). What is the Panel's view on the adequacy of this approach?

9. **Model Parameterization.** Four parameters were varied in each Monte Carlo analysis performed for a specific organism associated with a given scenario: the magnitude and shape of the exposure curve and the slope and intercept of the dose response curve. What parameters does the Panel believe should be varied in the lower tiers of a probabilistic risk assessment? For the case study, toxicity data from standard toxicity test protocols with a narrow range of animal age and size and test condition were used. What does the Panel believe with respect to the expression of generic effects ignoring size, age, feeding, respiration rate, etc.? Is the generic prediction approach sufficient or should the model include consideration of variations in these parameters?

10. **Routes of Exposure .** Due to the high solubility (~700 ppm) of ChemX in water, dietary and sediment associated routes of exposure were not considered. Does the Panel agree that this is sufficient for ChemX? What are the Panel's thoughts on when these additional routes should be considered, in terms of specific physico-chemical parameters and values?

Detailed Response to the Charge

General Comments

The report and progress on incorporating a probabilistic approach in lower levels ("Tier 2") toward assessing the risk of agricultural chemicals to aquatic organisms represent a good overall effort which should move the aquatic risk assessment of pesticides forward from the current deterministic approach. The Panel commends the Agency's progress in implementing this approach. It was obvious to the Panel that much effort and critical thinking has gone into this process and that progress has been amazingly fast. Although some details remain to be completed the proposed approach should allow for a generic probabilistic assessment to be carried out.

The Panel strongly agreed that the Agency has progressed a long ways since the last SAP meeting on this subject (April, 2000). It was uniformly concluded that the Agency has done a good job and is at the forefront of conducting an ecological probabilistic risk assessment (PRA). A discussion ensued over the value of the probabilistic approach: We may develop a sophisticated model, but of what value is it, if it cannot be implemented or used. A large question facing the Agency concerns how to make decisions using these results. In terms of the future, the Panel concluded that field verification (for effects and chemical fate) of model predictions is very important and needs to be conducted.

Also, in enhancing the scientific underpinnings of PRA, when making PRAs based on acute data, it will be helpful to look at the adequacy of the 96 hr tests to estimate safety of chemicals in the environment and to consider the uncertainty of post-exposure mortality.

The use of conservative inputs as opposed to realistic inputs needs careful consideration. Conservative results should be reserved for initial screening level assessments. If the registrant cannot move to more realistic inputs, then there is no need for different levels of assessment. For example, the maximum pond concentration over 36 years using data from the nearest weather station is used in the first level of assessment. This is a screening-type assumption. For higher levels of assessment, one might more properly use a distribution of calculated concentrations, not the peak, especially if running tens of thousands to millions of

replications. Perhaps Level 1 or Level 2 might use peak concentrations, with the next tier using a distribution of calculated rather than peak concentrations.

The Panel felt strongly that there should be an attempt to obtain comparative data for active ingredient and formulation, even at the "early" Tier 2 level. In situations in which aquatic exposure is through direct deposit or drift into water bodies, available evidence suggests that formulation may enhance pesticide toxicity. Ideally, formulation toxicity should be a requirement of registration. Where unavailable, there could be a compensating factor applied to the active ingredient.

In an effort to provide scientific strength to the PRA process, additional thought should be given to toxicity test methodology and analyses, including extrapolating from acute to chronic exposures when there are few chronic studies available. An example might include sensitivity analyses for the adequacy of using 96 h LC_{50}s without extrapolation for all "acute" exposure durations, e.g., 0, 1, 4 days of exposure. Exponential curve fitting has been suggested by Giesy (1994), Mayer *et al.* (1994), and Mayer (1999), but conventional models applied in other fields include lognormal, log-logistic, Weibull, gamma or several other models. Second, Suter (1998) discussed extrapolation from acute to chronic data, particularly the uncertainty associated with excluding post-exposure mortality in predicting effect levels from exposure data (Newman, 2001). A third issue in testing methodology is understanding the proportional response. This represents either the proportion of a field population dying or the probability of an individual dying during the exposure period (Newman, 2000).

1. Exposure Model Input Parameter Variability. In addressing the regional effects to the farm pond, we have used the exposure model matrix by looking at the combination of 3 pHs, 3 field-to-pond size ratios, and 2 soil aerobic metabolism rates, without changing the meteorological data. We are currently pursuing development of an exposure model to include the following parameters as variable inputs into PRZM/EXAMS: field/pond size, Kd, soil aerobic metabolism, application date, pond depth, and pH. What other parameters should be considered as variable distributions? Would the Panel please list other recommendations or suggestions for refining this approach, considering our purpose of PRA?

The Panel's consensus was that farm ponds should be viewed as surrogates for aquatic receiving-water systems. As such, it is aspects of water quality influences on pesticide exposure that need to be taken into account. For many compounds, chemical speciation needs to be considered as a factor that may control whether other variables need to be distributed. This is not important for Chem X, as it is not an organic acid or base. For many pesticides, acid-base speciation is a very important factor in determining fate and toxicity. Phenoxyacetic acids (2,4-D and

2,4,5-T) are good examples of this, where sorption (K_d) is a function of pH. Sorption of organic cations (protonated bases) also is pH dependent and cannot necessarily be normalized to soil organic carbon or even adequately to pH. Thus, it is important to consider the basic chemistry of the compound under investigation and for those compounds, pH should be treated as a distributed random variable. Sorption, hydrolysis, toxicity, as well as other processes will be pH dependent. Thus, it is important to consider that pH can affect chemical toxicity either directly (ammonia is toxic, whereas ammonium is much less toxic) or indirectly (hydrolysis rates).

For a chemical with high water solubility and low abiotic decay (photolysis and hydrolysis), microbial decay within the water body may be the greatest elimination process, such that subsequent runoff events would not lead to buildup in concentration in the water body.

Other parameters the Panel recommended for consideration:

1. **Number of events and application date**. As a minimum, more total and precipitation events need to be considered. This can be accomplished by using rain gauge information across a region and "date of application" as either a variable or distributed variable. A data set that includes 20 regional rain gauges with 36 years of data, with a possibility of 30 different application dates leads to over 20,000 total events. Even if only two runoff events occurred per month, a database of over 1,000 runoff events would result.

2. Soil Aerobic Metabolism. Rather than make this parameter independently distributed, it is likely that it could be made a simple bounded function of soil moisture (calculated with PRZM) and temperature.

3. Pond Depth. Unless photolysis and sorption to sedimentary materials are important, the depth of the pond is not as important as the volume of the pond, specifically the area:volume (a key variable). The volume is important due to dilution and retention.

In the problem definition section of the assessment, there should be a consideration of the ecological value of the aquatic resources to be protected. Such a consideration would have concluded, for example, that a key aquatic resource of the midwest area being modeled is the ephemeral or semi-permanent ponds, many of which are within crop areas proper. The invertebrate biomass of semi-permanent ponds are critical to waterfowl production. Therefore, the scenario of a permanent water supply used here may fail to protect one of the most important biological resources in the target area. The Agency should look at Sheehan et al. (1995) for more details.

4. pH. If pH does not affect any of the fate processes, then the output will be insensitive to pH. In the case of ChemX, hydrolysis is pH dependent and this dependence can be determined precisely as a function of pH to within 5 to 10% error. The question therefore really is: What is the variability of pH in farm ponds in the region of application (can be seasonal, affected by liming of soils, level of eutrification)? In addition, what are pH values of soil pore water? Because hydrolysis is an exponential function of pH, more is gained in the analysis if results at different pH values are evaluated. In the analysis of ChemX, however, using pH as a distributed parameter may have resulted in questionable results due to the lack of knowledge regarding the true hydrolysis rate constants.

5. K_d. Regarding sorption, it is recommended that fraction organic carbon rather than K_{oc} be used as a distributed parameter in the calculation of K_d. The fraction organic carbon is available in several large databases and can vary greatly regionally, whereas K_{oc} for most compounds is relatively constant (factor of 2).

Whether other variables should be distributed may require some additional scientific expert analysis as to whether they will influence the outcome. Toxicity generally is a logarithmic function of exposure concentration. Hence, any parameter that affects exposure concentration linearly, yet is logarithmically distributed, surely needs to be considered. Many variables however are not logarithmically distributed. However, the combined effect of a factor of two-to four-fold variation in these variables may be significant, especially when examining the most sensitive species. In effect, all chemical fate parameters need to be given initial consideration. These include, hydrolysis, photolysis, volatilization, soil aerobic metabolism, water column metabolism, and soil-water partition coefficient. Most of these processes can be assigned pseudo-first order reactions with associated uncertainty based on the literature or chemical structural information. Hence, data or technical opinion needs to exist to eliminate pathways from consideration. A question is raised: which and when do these process coefficients exist as linear functions (i.e., $k_{observed} = k_{hydrolysis} + k_{biolysis}$), such that some values do not need to be considered further due to their relative magnitude in comparison to other parameters. This may be the case for soil aerobic metabolism rate in relationship to hydrolysis, which appears to have no effect on the outcome (Agency background document, page 58, figure 9) for ChemX. Resources can be saved by excluding these unimportant transformation pathways from full probabilistic analysis or from further analyses altogether.

Site specific variability: K_d (minor), metabolism (minor), pH (minor), precipitation (major), field/pond size and pond depth (minor) – processes that are a function of temporal conditions have site specific variability.

Regional variability: K_d (minor to major depending on variation in soil organic carbon within the region), metabolism (minor to major), pH (major, for example from pH = 6 to 10), precipitation (major), field/pond size (major) – processes that are a function of temporal *and* spatial conditions have regional variability.

Input parameters to PRZM – If application date is considered, this is related to runoff timing and volume and concentration within the runoff, yet there are other PRZM input parameters that also effect volume and chemical concentration in the run-off (such as slope and soil type) which are used in the Soil Conservation Service curve number method.

2. Exposure Distribution Profile Selection. The exposure component of the aquatic risk assessment model uses 36 year rainfall data to generate 36 annual maxima for exposure concentrations. Two approaches were employed in establishing an exposure distribution profile: a theoretical fitted distribution using Monte Carlo analysis and an empirical distribution using a bootstrap method. Both methods performed similarly except in the tails of the distributions. The empirical distribution is preferred due to its objectivity and speed of calculation. What should the criteria be for choosing between theoretical and empirical methods?

Given the available data set used by the Agency, the Panel agreed that the establishment of distributions was properly performed. Because there appears to be no consistent theoretical function, in the sense that different data sets are best fit by different functional forms, the strong Panel consensus was that empirical data are preferred. The published literature indicates that a bootstrap method is widely used. Given the nature of limits on the number of iterations (re-sampling), the empirical distribution using a bootstrap method and would seem to be a good criterion. The Panel discussed two major issues: 1) How different are the distributions? If they are not different, then practical considerations should apply and 2). If the distributions are different, which is the critical one? Quantile/quantile plots could be used to determine how they differ.

Both distributions are continuous estimated theoretical distributions, so that one can compare a variety of quantiles, including quantiles in the extreme tails. The comparison of distributions should not be restricted to the 36 observed values. It was suggested the data could be summarized into Kolmogorov-Smirnov, Anderson-Darling, or Kullback-Leibler distances, compare moments, or compare results. This is a comparison of populations and not a hypothesis test. In all cases, there is a need to compare observed distance or difference to *a-priori* choice of what constitutes a 'large' difference.

The Panel acknowledged that a very difficult question facing the Agency is

"Which distribution (if they are different) is closer to the true distribution? Specifically, is a 20% difference important?" One could compare biological results but do the distributions have meaning? If both are different, there was agreement that the one most realistic should be used.

One Panel member provided an example algorithm to cross-validate the models and determine which would give information about behavior of the models in the tail regions (≤ 0.05). The suggestion is to calculate a log likelihood of the point differences, summed over all points.

Define F(x) as the estimated cumulative density function (CDF) using the empirical distribution and G(x) as the estimated CDF using the theoretical distribution. Each defines a probability density function, f(x) and g(x). Define $F_{-i}(x)$ as the estimated CDF using the empirical distribution when the ith observation is omitted and $G_{-i}(x)$ is the estimated CDF using the theoretical distribution when the ith observation is omitted. Each defines a probability density function, $f_{-i}(x)$ and $g_{-i}(x)$. The log-likelihoods for the ith observation are $\log f_{-i}(x_i)$ and $\log g_{-i}(x_i)$. Each log-likelihood is close to zero when the observed value is very likely (high probability) given the model and small (very negative). The cross-validation statistics for the empirical and theoretical models are: Sum (over i) $-\log f_{-i}(x_i)$ and Sum (over i) $-\log g_{-i}(x_i)$. The model with the smaller sum is the model with a better fit by the data.

To capture the full spectrum of possible rainfall events, the upper bound of rainfall events, if not all rainfall events, should be generated by using regional data, e.g., data obtained from the appropriate corn growing region and not simply from a given site. The suggestion was made to obtain the maximum rainfall during desired time intervals, apply that to the field on the day of pesticide application, and use the resultant runoff to estimate the upper bound of runoff. The Panel concluded that the entire regional weather data set should be employed to better estimate 50- or 100-year rainfall events.

However, data are also needed to define the more common rainfall events, as protection provided by selecting the 50 or 100 year events may be minimal given 1) the frequency of occurrence of smaller rains still leading to erosion, and 2) the practical ability of any xenobiotic to successfully avoid real or "apparent toxicity" due to its presence in runoff (sediment loading may cause stresses (e.g., oxygen depletion) that may be sufficient to produce direct mortality).

There are deficiencies within the rainfall data set. The data describe only daily totals. Intense rainfall over short time intervals needs to be accounted for in modeling processes at higher tiers. Given that soil infiltration and evaporative transports are important parameters in PRZM, the time interval of rainfall must also

be an important variable. This refinement also requires runoff modeling at shorter time intervals.

Finally, an analysis should be carried out to check whether there has been a systematic change in the frequency or severity of extreme weather events over time as would be predicted by many global warming models.

3. Interspecies Variability. In developing a sensitivity curve for freshwater fish with the available data, species' data were combined into their respective families. This was done because all families except salmonids had a single representative species, whereas, salmonids had four representatives. The aim was not to skew the sensitivity data by the over-representation with salmonids. The geometric mean of the multiple species and/or multiple tests with the same species was used in establishing points along this curve. What does the Panel think of this approach? Please provide alternative recommendations, if any, for dealing with limited data sets.

To understand the consequences of using the geometric mean to describe dose-response curves, consider multiple tests with different estimates of the LC_{50} and its slope. There are at least two explanations: 1) sampling uncertainty, quantified by the standard error of the estimates. If this is the cause of the differences, using the geometric mean provides the best estimate of the dose-response curve and 2) variation due to organism size, age, test water conditions, and so forth. If this is the case, using the geometric mean introduces unquantified uncertainty to the dose-response curve. The 'true' dose-response curve could be one of the measured dose-response curves (if that test were conducted in conditions that most closely match the field conditions) or a mixture of the measured dose-response curves (if conditions varied between those for the various tests). Use of the geometric mean is reasonable in the absence of any additional information because effect is proportional to log of concentration, not concentration.

In general, it appeared advisable to the Panel that the Agency should assess species' sensitivities in such a way as to prevent one phylogenetic group from skewing the distribution, either due to sensitivity or insensitivity. However, when toxicity response for a given phylogenetic group is spread across the entire data distribution, there is no compelling reason to calculate genus mean values. In the case where salmonids or cladocerans, for example, are clearly clustered, there may be a good reason to use genus mean acute values. There are a few rules of convention that might provide guidance: (1) one should not generate family mean acute or chronic values, i.e., reduction to the family level cuts across a large number of different genera with different life stages and sensitivities; (2) care should be exercised in calculating genus mean acute values when threatened or endangered species are being considered or when there is reason to believe the species

differences are real; (3) genus mean acute values are useful when given toxicity values for the same species or genus differ greatly with no apparent explanation; and (4) calculation of species mean acute values is routinely used unless the values are known to be representative of a given environment and are not a function of laboratory or method variability.

There was general agreement with using a representative of each family rather than including all species separately. Otherwise, large variability among species may lead to results from a few families dominating the sensitivity distribution. Results of multiple tests on a given species can vary by an order of magnitude or more. Therefore, the geometric mean is a reasonable approach for choosing one value for the species. One could always use the arithmetic mean. However, for the data presented, the result would be a higher value than for the geometric mean (and thus less protective). The Panel discussion suggested that choosing between the two may be a philosophic distinction more than a scientific distinction.

Ultimately, the only alternative to limited data sets is to obtain more data sets. The Panel discussed, at some length, the general concern about representation of the taxocene, guild, community, or sensitive group about which one is making a statement of effect or risk. The broader question of representativeness must always be entertained, as well as the consideration of what is being protected (e.g,. trout, farm ponds, and freshwater fish).

4. Effects Input Distribution. The extrapolation of fish sensitivities used a lognormal distribution of toxicity (LC50s) and a normal distribution of the dose-response slopes. What does the Panel think of this approach and which other approaches could have been used?

The Panel discussed at some length the appropriateness of using the lognormal to precisely reflect species sensitivity distributions. One Panel member pointed out that Shapiro-Wilks tests of log normality for half of data sets randomly selected from the EPA ACQUIRE data set indicated that the log normal model is not generally valid (Jagoe and Newman,1997; Newman et al., 2000). Although a public commenter suggested that the lognormal distribution does fit if logical species' subsets are used for the analyses, the exploration of a series of candidate distributions for parametric models is probably better than assuming a log normal distribution. In cases where no specific model is acceptable, application of nonparametric methods such as a bootstrap could provide a partial solution. The above cited papers contain such examples. Russell Erickson and Charles Stephan (2001) of EPA recently criticized the particulars of the bootstrap methods in those manuscripts and provide good suggestions for improving the approach. The approach is similar to that discussed early for bootstrap "mixed empirical-exponential CDF characterization of uncertainty" by the Agency. To include both

variability and uncertainty in dose-response curves, first one could conduct a nonparametric bootstrap of the original toxicity data and estimate the slope, intercept, and their associated standard errors and correlation. Second, draw of estimates of the slope and intercept from the estimated distribution to use in the simulation is applied. Finally, repeat the whole process, beginning with a nonparametric bootstrap sample from the data (a continuation of the discussion concerning Question 2).

It would be useful to include the actual data used in fitting the lognormal models (estimates in Table 6). Doing so would make it possible to try fitting other model forms. Also, it should be understood where the concentrations and toxicity responses of the studies lie in comparison to estimates produced by simulations and with respect to the center or tails of the distribution. Doing so would help determine to what extent the lognormal is appropriate. Because sensitivity analyses show that model results are very sensitive to choice of slope and intercept of the concentration-response curves, it would make sense to fit other candidate curves to the actual data. Examples might be the log logistic or the log complementary log-log, all very easily conducted these days using SAS or Splus. Then candidate models could be re-analyzed to see to what extent results differ. These models do have different behavior in the tails ($\leq$ 10 % or $\geq$ 90 %) that might produce different results. This might be especially true when looking at the 5th percentile species. However, as the Agency has been using the lognormal for years and it has served as a standard, the Panel recommends that the Agency consider using models other than the lognormal in a more general context of revisiting toxicity studies. With current-generation software, it is very easy to fit other models to the data. If the models are in reasonable agreement, confidence can more easily be placed in the predictions.

Regarding the use of a lognormal distributional model to assess fish sensitivity, previous reviews have investigated the use of other distribution models and their impact on the 5th and 50th percentiles. That review leads to the conclusion that similar answers are obtained regardless of the distribution model used when the data sets are fairly large (Toll et al., 2001). This assumes that the distributions used have been shown to fit the data. Greater concern exists for model selection when small data sets are used. General guidance in selecting a model is that the model should appear to fit the data especially in the tails or region of interest.

Appendix F of the Agency's background document should be labeled "Estimates of Toxicity used in the Monte Carlo Simulations" rather than "Data." Also, "simulation output" is not data. It is either simulation output or, possibly, simulated data. Care should be taken in the use of the terms data, estimates, and simulation output or simulated data. Sensitivity analyses in Tables 44 *et seq* include a 'correlations off' scenario. That scenario is problematic. Setting the correlations

to zero makes the distributions dependent on the units of concentration. Setting the correlation to 0 and expressing concentration in ppm will give different distributions than expressing concentration in ppb. When the empirical correlation is used, the effects distributions are invariant to the scaling of dose.

5. Extrapolation with Limited Data. Since only one acceptable toxicity test was available for an aquatic invertebrate, an extrapolation using toxicity profiles of other compounds in the same pesticide family to determine the average sensitivity of the tested species and extrapolate a species sensitivity distribution was employed. What does the Panel think of this approach? What are the Panel members opinions on alternative approaches that may be used for establishing a sensitivity profile across diverse invertebrate taxa when one or a very few tests are available?

The Panel was uncomfortable with the relative lack of data. The recommendation was to see if sufficient data are available for Monte Carlo determinations. As this discussion dealt primarily with Tier II, the minimum number of data points for any kind of confidence is critical. Although five may be the minium needed to represent species sensitivities, there was no agreement over what is "sufficient". A Panel recommendation is to re-visit this topic at a future SAP meeting and discuss specifically how to get sufficient data for a SSD.

The small gains in having raw data that allows for re-analysis are far outweighed by eliminating data that give valuable information on different taxa. Whatever model is fitted to the toxicity data will make a minimal difference in the estimate of average toxicity. It is therefore not productive to eliminate data that do not lend themselves to a re-analysis. Provided the data meet a modicum of scientific criteria, they should be accepted. A case by case analysis should be presented for data that are rejected.

Whereas the Panel would prefer to have additional invertebrate data on Chem X, the Agency's approach was creative and extended the data set. This is a difficult problem and the Panel lauds the Agency for having taken a strong approach with the limited data. Although this approach has limitations, it may be useful for a Tier II assessment where the opportunity exists to collect additional information in Tier III should the risk estimate warrant this effort. Supporting the current approach was that the substances had the same apparent mode of action and that *Daphnia magna* values were available for each of the substances, providing a common EC_{50} to use as the starting point for the distributional simulation. The increased number of invertebrate data points allows a distribution to be created for invertebrates and an overall expanded species sensitivity distribution for Chem X. The most sound approach would be to generate a sufficiently complete data package from which a species sensitivity profile can be constructed. It was the opinion of the Panel that

tests on a single species cannot adequately represent the diverse range of invertebrate taxa. Current core data requirements provide for tests on two birds, two fish, and multiple plants but only one aquatic invertebrate. In the current case, aquatic invertebrates appear to be the most sensitive aquatic group tested. In such cases, it is necessary to have a reasonable data set to undertake a sound assessment.

Unfortunately, lack of data is the norm rather than the exception in risk assessment. The single freshwater aquatic invertebrate and single estuarine invertebrate toxicity datum present a major problem. No reasonable distribution can be drawn for that group of organisms. To realistically perform these probabilistic models, more toxicity data are necessary. If only one datum exists, more data are necessary before probabilistic assessments can proceed. Extrapolation of a sensitivity distribution from data on similar chemicals is one possible approach; however, great caution must be taken to avoid statistical and mathematical gymnastics on too few data points; the results will probably be weak. The underlying assumption of this approach is that because the mode of toxicity is the same, the species sensitivity distributions would be similar for chemicals in the same class. There was no evidence that this assumption had been investigated.

In the case of Chem X, data for similar chemicals were available. Provided there is some indication that the underlying assumptions are reasonable, a case could be made for some sort of extrapolation, as has been done. In the case where there are no usable data for similar chemicals, one alternative may be the use of generic extrapolation factors to determine a theoretical value for a sensitive species (as was suggested for avian assessments).

There was a second assumption made in creating a model invertebrate concentration-response curve about which the Panel was not explicitly asked to address. It should be noted that to construct a concentration-response curve for "model" freshwater invertebrate species, the slope value used was from the marine species, pink shrimp, not for daphnia. Such extrapolation presents a considerable source of uncertainty. This is where caution should be exercised, as no justification was presented in the document to allow its use. This led to the use of limited species in PRA.

The focal species approach is used by ECOFRAM, although the extrapolation factor approach is more rigorous than the one described here (Mineau et al., 2001). The SAP encourages the Agency to adapt this method to aquatic assessments as well. This focal approach is distinguished from the 'safety factor' approach used in mammalian toxicology where arbitrary factors (of 10 usually) are applied to account for putative inter-taxa differences in susceptibility. Extrapolation factors are not arbitrary but are species-specific, empirically derived constants. The

extrapolation factor approach is a last resort when too few points are available to fit to a distribution. Again, this approach is second best to finding more toxicity data points and deriving a compound-specific SSD. The method, of course, is very error prone and could result in serious underprotection. For this reason, it is recommended that larger confidence bounds be used.

The extrapolation factor approach uses existing knowledge about relative interspecies sensitivity. In birds, it has been found that some species are typically more sensitive than others to pesticides. Another approach called universal safety factors, uses the same extrapolation factor for any starting species, assuming this one has been chosen at random (Luttik and Aldenberg, 1997). This is an option if, on the aquatic side, relative interspecies susceptibility patterns cannot be established.

The avian risk approach consists of obtaining as many large data sets as possible (e.g., data sets of six species or more). For each pesticide, an HC5 is obtained through the Aldenberg and Slob (1993) approach. The HC5 can be determined with any desired confidence bounds. For every species routinely tested, the mean and SE of the distance between that species' LC_{50} and the HC5 is computed. The mean distance becomes the extrapolation factor. Extrapolation factors can also be calculated for combinations of species by taking the geometric mean LC_{50}. A similar approach was suggested by a Panel member. Specifically, this entails resampling of the individual species' distances from the HC5 rather than making the implicit assumption that those distances conform to any preconceived distributions.) Generally, multiple species factors are more stable (e.g., lower coefficient of variation, CV) than single species factors. When there is a choice of which extrapolation factor to use to derive the HC5 for a new pesticide, the extrapolation factor with the lowest associated CV is used. On average, the error will not be as great. The confidence interval around each extrapolation factor can also be used to obtain higher confidence of protection (for birds, using the upper 95% confidence interval of the extrapolation factor is fairly equivalent to computing the 95% lower confidence bound of the HC5).

In terms of applying the approach to the current data set (or more specifically the current paucity of data), some problems are foreseen. First of all, preliminary analyses have not been conducted to determine whether separate taxa do follow some systematic differences in sensitivity across pesticides. This should be done to see whether a more robust extrapolation factor would be obtained by using data from a broader group of pesticides (i.e., more than four). Second, the regression model used by the Agency to estimate the approximate percentile sensitivity position of *Daphnia magna* for the four pesticides in the same chemical family as ChemX does not provide a correction factor for small samples as does the Aldenberg and Slob method. It would be advisable to estimate the HC5 for each of

the four pesticides using a method that corrects for the reduced ability to establish the distribution variance for small sample sizes. Once this is done, the distance between the HC5 and the *Daphnia magna* data point can be obtained for each of the four pesticides and used (e.g., via resampling) to estimate the Chem X HC5. Obviously, this analysis should be subject to simulations to determine error propagation. And, third, care needs to be taken when using species from freshwater or marine systems to estimate sensitivity distributions. For example, with triazines, the relative sensitivity of marine and freshwater organisms is probably constant. However, for an organophosphate, the relative relationships between fresh- and saltwater species is very likely different.

6. Taxa Aggregation. Freshwater taxa were separated from marine taxa in this case study. Since the marine data sets were limited to a single test species, toxicity profiles for that species were used in the assessment and no sensitivity distribution across taxa performed. Data from other related pesticides for marine species were also not available as in the case for freshwater invertebrate taxa. What is the Panel's opinion of separating toxicity data from marine organisms from data on freshwater organisms for purposes of establishing sensitivity profiles? Would the Panel please provide an alternate recommendation, if it has one?

Although there are no easy answers, the Panel did agree that it is reasonable to separate saltwater and freshwater data into two subsets. Such an approach may be the most reasonable action unless it can be demonstrated that the two data sets (fresh- and saltwater species) are comparable, in which case data could be pooled (Mayer and Ellersieck. 2001; Mayer, 1999; Mayer, et al., 1994; Mayer, and Ellersieck. 1986). Given the limited data sets, the need is apparent to try and use as much data as possible, arranged in a sensitivity distribution of all species (perhaps using ACQUIRE and studies a bit broader than used here).

It was noted by several Panel members that exposure is quite different between marine and freshwater systems. Marine pH is higher and hydrolysis rates are typically higher than in freshwater. It could be argued that there is justification for separating freshwater and marine taxa based solely on biological, and/or physiological differences. Thus, separating taxa should be the norm unless sufficient data and a biological basis exist to justify pooling the two groups. However, it should also be noted that fate studies are only done in freshwater systems. Thus, extension exposure modelling is really only done for freshwater systems. The potential for differences in chemical fate in freshwater and marine/estuarine systems may introduce considerable uncertainty into the assessment. As already noted by the Agency, more work is required on marine/estuarine exposure scenarios.

7. Chronic Assessment. Available chronic data were limited and, therefore, fewer scenarios were considered sufficient to cover the range of outcomes. An exceedence probability approach was taken to evaluate the potential of chronic effects. What alternative approaches to evaluating chronic effects could have been taken?

The Agency's ability to conduct a PRA using joint distributions is clearly limited by the lack of chronic effects data. One alternative approach is to generate a chronic distribution of no-observed effect levels (NOECs) using the acute data set and applying an acute to chronic ratio (ACR) to develop predicted chronic values. This is the same approach that the USEPA Office of Water uses to derive water quality criteria. The approach published by Stephan *et al.* (1985) derives an ACR by using three or more measured pairs of acute and chronic studies. The mean of these values is then used to convert the final acute value (95^{th} percentile/2) to a final chronic value.

For the existing aquatic data set, there are only two matched pairs of acute and chronic tests, rainbow trout and *Daphnia magna,* with respective ACRs of 12.33 and 2.96 (mean = 6.04). For purposes of this example, an ACR of 12.33 was chosen to convert other fish acute values to estimated chronic NOECs. This value was chosen as there were only two ACRs and because all estimated ACRs were for fish. Using the fish ACR appears to be the most appropriate and conservative. The example data set in Table 1 provides a distribution of values for assessing chronic toxicity. The 5^{th} and 95^{th} percentiles of the distribution were calculated, as noted in Figure 1, using the approach of Aldenberg and Slob (1993).

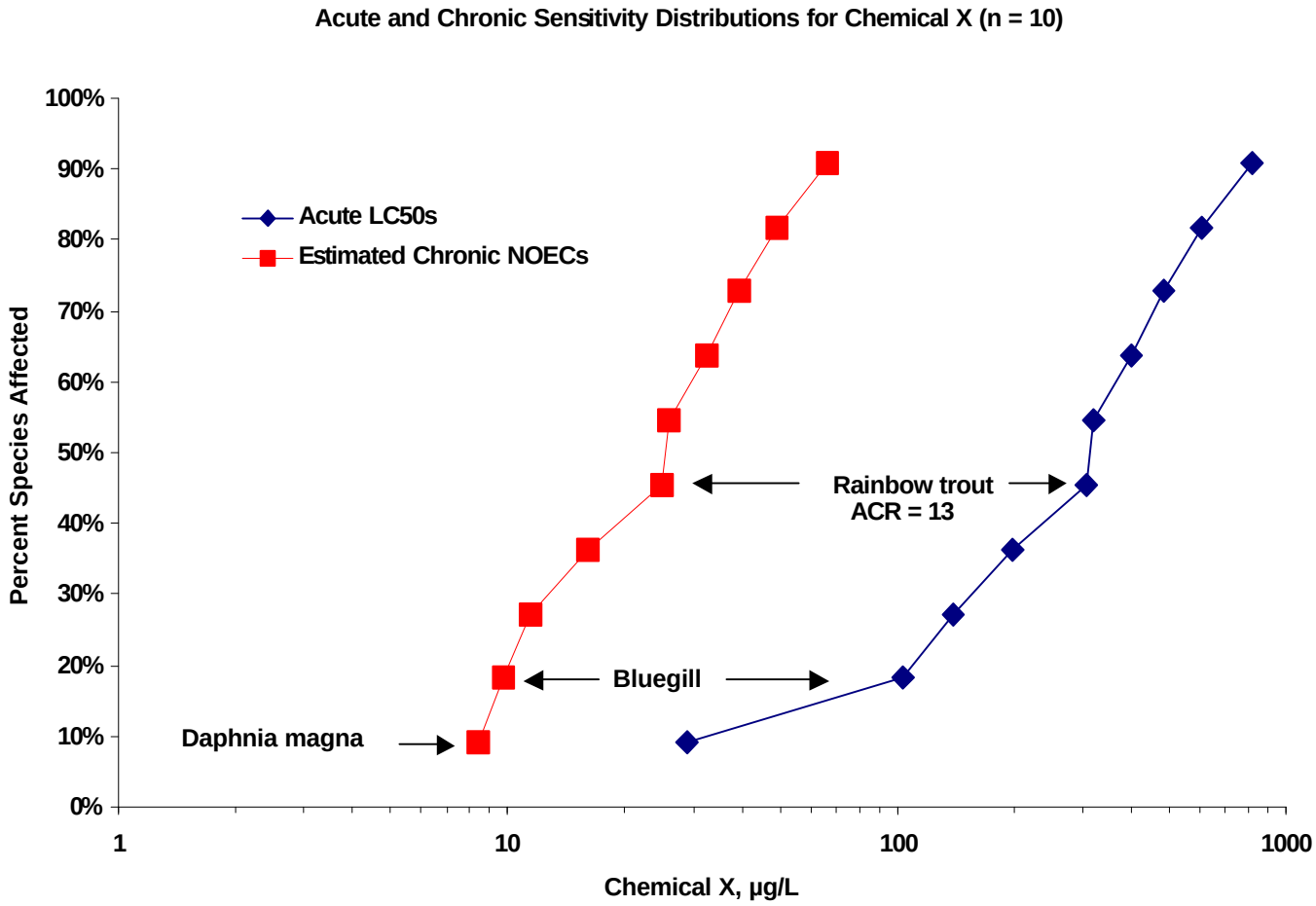
Whereas some uncertainty is introduced by this approach, the Panel concurs it is better to develop a sensitivity curve than to simply use one or two chronic data points for the assessment, which clearly results in a large degree of uncertainty associated with the level of protection being provided.

Table 1: Chem X-Chronic Based On Mean ACR

Species	LC50 (ug/L)	ACR	Estimated NOEC (ug/L)
Fathead minnow	821	12.33	66.6
Steelhead	606	12.33	49.1
Coho salmon	486	12.33	39.4
Brown trout	402	12.33	32.6
Channel catfish	320	12.33	25.9
Rainbow trout	306	12.33	24.8[a]
Yellow perch	198	12.33	16.1
Lake Trout	140	12.33	11.4
Bluegill	104	12.33	8.43
Daphnia magna	29	2.96	9.8[a]
Estimated chronic 5th percentile			6.77

[a] Measured NOEC

Figure 1.


Acute and Chronic Sensitivity Distributions for Chemical X (n = 10)
Percent Species Affected
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
10
100
1000
Chemical X, µg/L
Acute LC50s
Estimated Chronic NOECs
Rainbow trout
ACR = 13
Bluegill
Daphnia magna

The Panel concluded that, given the lack of acute data available, coupled with a lack of a chronic concentration-response for many studies, the "frequency-of-exceedence" approach is reasonable at a Level 2 assessment. The information could be used to trigger a more thorough investigation of chronic effects for the group in question at a higher level of refinement. In the current case of Chem X, an ACR for rainbow trout was used to determine a expected no-effect concentration (NOEC) for bluegill. Some consideration should be made of what the variability and uncertainty of the ACR is in this case. This might include looking at mean values for other products in the same chemical class.

8. Estimation of Species Sensitivity. In this case study, probabilistic assessments were performed for specific species (e.g., bluegill, sunfish, and rainbow trout) and for extrapolated species (e.g., 5th percentile sensitive and 50th percentile sensitive). What is the Panel's view on the adequacy of this approach?

Part of the purpose of a PRA is to understand the range of possible effects. With that in mind, the Agency's approach seems reasonable. In the context of aiding understanding, including results for a couple of species that traditionally have been evaluated is also reasonable. This concept provides a means of estimating the risk to the generic 5th percentile and to specific fish species of interest. Assessment of risk at the 5th percentile allows an assessment for sensitive species within the overall distribution, whereas the specific risk assessments for bluegill or trout are useful for given site-specific applications (farm ponds, small streams). However, the Agency could give further consideration to using the entire species sensitivity distribution when calculating aquatic risk as opposed to separating the distributions between invertebrates and fish. The methodology is very sensitive to sample size, and separating the data into groups (fish and invertebrates) reduces the size of each data set and introduces unnecessary uncertainty into the risk estimates due to the small sample sizes in the distribution.

For purposes of illustration, the 5th percentiles of the species sensitivity distributions are calculated with the fish and invertebrate data combined in a variety of ways. The overall methodology is sufficiently simple that risk comparisons could be calculated with and without combining the distributions. The use of the (generic) 5th percentile (HC5) is frequently used in conducting aquatic risk assessments because field data exist which conclude that, for most substances, the HC5 provides a sufficiently low concentration as to be protective of organisms in natural systems. Versteeg et al. (1999) demonstrated this for 12 substances.

9. Model Parameterization. Four parameters were varied in each Monte Carlo analysis performed for a specific organism associated with a given scenario: the magnitude and shape of the exposure curve and the slope and

intercept of the dose response curve. What parameters does the Panel believe should be varied in the lower tiers of a probabilistic risk assessment? For the case study, toxicity data from standard toxicity test protocols with a narrow range of animal age and size and test condition were used. What does the Panel believe with respect to the expression of generic effects ignoring size, age, feeding, respiration rate, etc.? Is the generic prediction approach sufficient or should the model include consideration of variations in these parameters?

The Agency's 'Guiding Principles for Monte Carlo Analysis' (p 16, point 10) indicates 'There are limits to the assessor's ability to account for and characterize all sources of uncertainty. The analyst should identify areas of uncertainty and include them in the analysis, either quantitatively or qualitatively.' The analysis of parameter sensitivity (e.g., Table 44, p 47) indicates that model output is sensitive to the effects parameter. Because the estimated slope and intercept of the concentration-response curve are highly correlated, sensitivity to slope may simply reflect a sensitivity to the LC_{50}. One role of the Tier 2 analysis is to guide data collection for subsequent analyses. If sources of uncertainty are ignored, the analysis provides no guidance for further data collection. Qualitative consideration (e.g. by scenarios) would give an indication of the sensitivity of the PRA to specific forms of uncertainty.

The Panel's judgment is that these additional sources of uncertainty should be incorporated in a Tier 2 PRA only if they can be quantified (explicitly or qualitatively). Otherwise, they should be deferred to higher levels of refinement. No matter which types of uncertainty are included, the presentation of the PRA should include a table where the uncertainties that are included in the analysis and a table of those that are not. The Panel concurred that there is no need to spend much time at this level on new tests or developmental tests. Rather, there are more key components for this Tier 2, such as latent mortality (time dependent) and sources of uncertainty. This tier is a screening-level assessment, so that one should vary as many input variables as there is information to indicate that variations are possible or reasonable. Experience in stochastic modeling dictates that only a few variables dominate variance in model output. For example, peak water concentration and a few risk parameters are very important in this model. These few variables would be the ones that receive the most scrutiny and additional data work if the PRA were to move to the next higher level of assessment.

In any assessment, there is a need to ensure that the effects data are somewhat representative of the exposure scenarios modelled (or vice versa). For example, in the case of Chem X, exposure scenarios modelled include ponds at three different pH values, yet, the effects studies are conducted within a standard pH range. Use of these effects data will introduce uncertainty into the assessment

and needs to be acknowledged. It would be appropriate to address this type of uncertainty at the next level of refinement, especially if the variable in question affects the overall exposure or effects outcome.

10. Routes of Exposure. Due to the high solubility (~700 ppm) of ChemX in water, dietary and sediment associated routes of exposure were not considered. Does the Panel agree that this is sufficient for ChemX? What are the Panel's thoughts on when these additional routes should be considered, in terms of specific physico-chemical parameters and values?

Consideration of water solubility alone is probably not sufficient. For example, if Chem X were a cationic organic, it is possible that it would have a high affinity for suspended clays and sediment. In addition, it may be important to consider the mode of toxicological action. The primary routes of exposure should be considered. Hence, if a compound is highly water soluble, the primary route for FW will be via gill uptake, not via diet. However, should compounds adsorb readily to materials in the water or sediments, then this must be taken into account. It should be noted that in those cases where partitioning to sediments occurs, effects data on suitable benthic organisms will also be required to conduct an appropriate assessment. ECOFRAM has suggested using the K_{ow} or K_{oc} of the chemical to determine whether to include these routes of exposure. In the case being evaluated by the Panel, the partitioning constant is not high enough to warrant evaluation of ingestion routes. The Agency should evaluate the partitioning of each chemical being assessed and then evaluate the need for ingestion or sediment toxicology. For future use in PRA, the Society of Toxicology and Chemistry is pursuing developing a workshop on dietary uptake. This would entail scientists from government, academia and industry to review and summarize the state-of-the-science on this topic.

REFERENCES

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

Erickson, E. and C. Stephan. 2001. Comment On: "Applying Species-sensitivity Distributions in Ecological Risk Assessment: Assumptions of Distribution Type and Sufficient Numbers of Species" (Newman et al., 2000). SETAC Globe May 2001.

Jagoe, R. and M.C. Newman. 1997. Bootstrap estimation of community NOEC values. Ecotoxicology 6: 293-306.

Luttik, R. and T. Aldenberg. 1997. Extrapolation factors for small samples of

pesticide toxicity data: Special focus on LD50 values for birds and mammals. Environ. Toxicol. Chem. 16:1785-1788.

Mayer, F.L. 1999. EPA/600/R-98/152.

Mayer, F.L. et al. 1994. Environ. Toxicol. Chem. 13:671-678

Mayer, F.L., Jr and Ellersieck, M.R. 2001. Inter-taxa Correlations for Static acute Toxicity to Aquatic Organisms. U.S. EPA., National Health and Environmental Effects Research Laboratory. Gulf Breeze, FL.

Mayer, FL and M.R. Ellersieck. 1986. Manual on acute toxicity: interpretation and data base for 410 chemicals and 66 species of freshwater animals. U.S. Fish and Wildlife Service, Resource Publication 160. 579 pp.

Mineau, P., A. Baril, B.T. Collins , J. Duffe, G. Joerman, and R. Luttik. 2001. Reference values for comparing the acute toxicity of pesticides to birds. Reviews of Environmental Contamination and Toxicology 170:13-74.

Newman, M.C. 2001. Population Ecotoxicology. John Wiley & Sons, Chichester, UK.

Newman, M.C., D.R. Ownby, L.C.A. Mezin, D.C. Powell, T.R.L. Christensen, S.B. Lerberg, and B.A. Anderson. 2000. Applying species sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient numbers of species. Environ. Toxicol. Chem.19:508-515.

Schafer, E.W. Jr., W.A. Bowles Jr. and J. Hurlbut. 1983. The acute oral toxicity, repellency and hazard potential of 998 chemicals to one or more species of wild and domestic birds. Arch. Environm. Contam. Toxicol. 12:355-382.

Sheehan, P., A. Baril, P. Mineau and D. Paustenbach. 1995. Predicting the effects of insecticides on aquatic systems and the waterfowl that use them. In: G.M. Rand. (Ed.) Fundamentals of Aquatic Toxicology (Second Edition), Taylor and Francis, North Palm Beach, Florida, pp. 827-857.

Stephan, C.E., Mount D.I., Hansen D.J., Gentile J.H., Chapman G.A., and Brungs W.A. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. U.S. EPA, Environmental Research Laboratory, Duluth, MN. NTIS No. PB85-227049. 98 pp.

Suter GW. 1998. Chapter 8 in Risk Assessment. Logic and Measurement, M.C. Newman and C.L. Strojan, eds., CRC/Ann Arbor Press, Boca Raton, FL.

Toll J, W. Adams, K. Brix, M. Burger, R. Cardwell, D. DeForest, L. Tear. 2001. Proposed approach for deriving predicted no-effect concentrations for data-rich substances: protecting aquatic ecosystems. Prepared for DETR (Department of the Environment, Transport and the Regions, Special Technical Meeting: PNEC Derivation for Data-Rich Substances, London, January, 2001.

Versteeg, DJ, SE Belanger, and GJ Carr (1999). Understanding single-species and model ecosystem sensitivity: Data-based comparison. Environ Toxicol Chem 18(6): 1329-1246.