SAP Report No. 99-03
May 25, 1999

REPORT

FIFRA Scientific Advisory Panel Meeting, May 25-27, 1999, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Session I **- Office of Pesticide Programs Policy for the Use of the FQPA 10x Safety Factor**

Session II **- Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment**

Session III **- Use of Watershed-derived Percent Crop Areas as a Refinement Tool in FQPA Drinking Water Exposure Assessments for Tolerance Reassessment**

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

TABLE OF CONTENTS

SAP Report No. 99-03A, May 25, 1999

REPORT:

FIFRA Scientific Advisory Panel Meeting, May 25, 1999, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Office of Pesticide Programs Policy for the Use of the FQPA 10x Safety Factor

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

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

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
May 25, 1999

Session I: Office of Pesticide Programs Policy for the Use of the FQPA 10x Safety Factor

PARTICIPANTS

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

FIFRA Scientific Advisory Panel
Ernest E. McConnell, DVM, Toxpath, Inc., Raleigh, NC
Herb Needleman, M.D. , Professor of Psychiatry and Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA
Christopher Portier, Ph.D, National Institute of Environmental Health Sciences, Research Triangle Park, NC
Mary Anna Thrall, DVM, Professor, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members
John Adgate, Ph.D. Assistant Professor, University of Minnesota, Minneapolis, MN
Richard J. Bull, Ph.D., Senior Staff Scientist, Battelle, Richland, WA
Jeanne Harry, Ph.D. National Institute of Environmental Health Sciences, Research Triangle Park, NC
Dale Hattis, Ph.D. Professor, Clark University, Worcester, MA
Ron Hood, Ph.D. Professor, University of Alabama, Tuscalossa, AL
Timothy Meredith, M.D., Professor of Medicine and Pathology, Vanderbilt University Nashville TN
J. Routt Reigart, M.D., Professor, Medical University of South Carolina, Charleston, SC
Patricia M. Rodier, Ph.D. Professor, University of Rochester, Rochester, NY
Peter Thomas, Ph.D. Director of Toxicology, Covance, Madison, WI
John Wargo, Ph.D. Associate Professor, Yale University, New Haven, CN
Lauren Zeise, Ph.D. California EPA, Berkeley, CA

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

PUBLIC COMMENTERS

Oral statements were received from:
John McCarthy, Ph.D. (American Crop Protection Association)
David Wallinga, M.D. (Natural Resources Defense Council)
Ms. Nancy Doerrer (American Industrial Health Council)
Ms. Lisa Lefferts (Consumers Union; Mothers and Others for a Livable Planet)
Mr. Todd Hepple (Environmental Working Group)
Richard Becker, Ph.D. (Chemical Manufacturers Association)
Mr. Eric Wilson (People for the Ethical Treatment of Animals)

Written statements were received from:
American Crop Protection Association
People for the Ethical Treatment of Animals

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency regarding Office of Pesticide Programs Policy for the use of the FQPA 10x Safety Factor. Advance public notice of the meeting was published in the Federal Register on May 5, 1999. The review was conducted in an open Panel meeting held in Arlington, VA, on May 25, 1999. The meeting was chaired by Ronald J. Kendall, Ph.D, The Institute of Environmental and Human Health, Texas Tech University/Texas Tech University Health Sciences Center, Lubbock, Texas. Mr. Larry Dorsey, SAP Executive Secretary, served as the Designated Federal Official.

The EPA Office of Pesticide programs (OPP) presented its policy for implementation of the FQPA 10x safety factor. As background to development of the policy, EPA convened a FQPA 10x Task Force to address toxicology and exposure considerations. The Task Force was charged with determining the appropriate set of child-specific exposure and toxicity information needed for decision making under FQPA and to review the Agency's tolerance decisions to ensure transparency, adequacy of documentation and consistency of Agency decisions with policy directives. Ms. Susan Makris (OPP/EPA) provided history and background of OPP FQPA 10x implementation efforts, Carole Kimmel, Ph.D. (National Center for Environmental Assessment, Office of Research and Development, EPA) discussed the 10x Task Force Toxicology Working Group Report (Toxicology Data Requirements for Assessing Risks of Pesticide Exposure to Children's Health), Linda Sheldon, Ph.D. (National Exposure Research Laboratory, Office of Research and Development, EPA) summarized the 10x Task Force Exposure Working Group Report (Exposure Data Requirements for Assessing Risks of Pesticide Exposure to Children), Penelope Fenner-Crisp, Ph.D. (OPP/EPA) described the policies employed by the Office of Pesticide Programs in making a determination regarding the FQPA Safety Factor when developing aggregate risk assessment and regulatory decisions for single active ingredient

pesticides and Ms. Brenda Tarplee (OPP/EPA) presented the revised OPP FQPA Safety Factor Committee Standard Operating Procedures.

The EPA, Office of Pollution Prevention and Toxics (OPPT) also presented a proposal for the minimum core toxicology database needed for the evaluation of chemicals for effects on infants and children, under the Children's Health Testing Program. As part of Vice President Gore's Chemical Right-to-Know Initiative, the OPPT is implementing the Chemical Right-to-Know Program. One aspect of this program is the Children's Health Testing Program, under which decisions must be made regarding the appropriate chemicals to test and the appropriate toxicology studies to be conducted. OPPT, in its implementation of the Children's Health Testing Program, plans to focus on chemicals to which children may have high potential exposure. The exposure characteristics of such chemicals might include those with high release to the environment and/or high exposure due to their presence in consumer products. Due to the potential for high exposure, many of the same considerations made for conventional food use products may apply for OPPT's Children's Health Testing Program. Jennifer Seed, Ph.D. (OPPT/EPA) summarized OPPT's proposal for the toxicology studies to be included in the core database for the Children's Health Testing Program.

CHARGE

The specific issues to be addressed by the Panel are keyed to the background documents *Toxicology Data Requirements for Assessing Risks of Pesticide Exposure to Childrens' Health*, *the Office of Pesticide Programs' Policy on Determination of Appropriate FQPA Safety Factor(s) for use in the Tolerance-Setting Process*, *Standard Operating Procedures for the Health Effects Division FQPA Safety Factor Committee* and *the Office of Pollution Prevention and Toxics Proposed Test Battery for the Children's Health Testing Program* and are presented as follows:

FQPA Safety Factor Issues

(1) Is a weight-of-the-evidence approach to making FQPA Safety Factor decisions appropriate, taking into consideration the toxicology and exposure databases for a pesticide and the potential risks for the developing fetus, infant and child as well as other populations? If not, why not? Given the scope of the evidence which OPP intends to consider, are there any other types of scientific information that OPP should consider in its FQPA Safety Factor determinations?

(2) Under what circumstances, if any, do you believe that OPP’s current approaches (the combination of empirical data, models, and assumptions) fail to yield risk assessments which are sufficiently conservative and do not understate the risks to infants and children?

Toxicology Database Issues

(3) Please comment on OPP's proposed criteria for defining the core toxicology database and its

approach to imposing a database uncertainty factor if certain key studies are missing from the database.

(4) After having considered the recommendations from this Panel and the Toxicology Working Group, OPP is beginning the process of calling in data for three studies (the acute and subchronic neurotoxicity studies in adult mammals and the developmental neurotoxicity study) for a subset of conventional chemistry food-use pesticides–known neurotoxicants. In addition, OPP will be proposing to require the same set of studies for all conventional chemistry food-use pesticides in the revision of the Part 158 regulations. Please comment on this two-stage approach.

(5) The OPP Policy Guidance indicates that one of the critical issues is whether or not to apply a FQPA Safety Factor pending receipt of newly-required studies. In the absence of the results from any of the studies to be required through data call-in notices (i.e., the acute and subchronic neurotoxicity studies in adult mammals and the developmental neurotoxicity study) what information from existing studies on a specific chemical would increase or decrease your concerns about the potential for pre- and post-natal hazard, in general, and for neurotoxicity and developmental neurotoxicity, in particular? Which, if any of the seven criteria discussed in section V.A.1.a., footnote 4 and associated text of the OPP Guidance are appropriate for judging whether there is increased concern about the potential for a pesticide to cause developmental neurotoxicity?

(6) Please comment on whether you expect that the NOAELs that are identified in the developmental neurotoxicity studies would, for a substantial number of chemicals, be lower than those NOAELs identified in the suite of studies historically required and used for age-related comparisons and Reference Dose derivation (e.g., prenatal developmental toxicity or multigeneration reproduction study, subchronic and chronic studies, etc.). Please explain the basis of your opinion.

(7) OPP is proposing to adopt the framework and its criteria/factors for assessing the degree of concern about the potential for pre- and post-natal effects as recommended by the Toxicology Working Group. Please comment on the appropriateness of the proposed criteria/factors for use in this assessment process, and OPP's proposed approach for accommodating its concerns in the Reference Dose derivation and FQPA Safety Factor decision processes, in the near term, and in the longer term. What scientific considerations relate to the addition of a safety factor where the hazard to infants and children is well characterized, and the data show that infants and/or children are more susceptible than adults?

Exposure Issues

(8) Subject to the qualifications expressed in the OPP Policy document and the report from the Exposure Working Group, OPP believes that each of the tiers for estimating exposure to pesticides through food, in almost all instances, will not underestimate exposure to infants and children. Please comment on this conclusion, as it applies to each of the tiers.

(9) OPP is developing a tiered approach to assessing the likelihood and magnitude of contamination of drinking water and its sources by pesticides. The Panel has been asked to comment on aspects of this activity at previous meetings. As an interim approach when direct assessment is not possible, is it reasonable and protective to regard the estimates generated by OPP's current methodology as upper bound pesticide concentrations for surface and ground water and to assume that this concentration will be found in drinking water?

(10) OPP is developing approaches to assess the likelihood and magnitude of exposure to pesticides in residential and other non-occupational use scenarios. The Panel has been asked to comment on aspects of this activity at previous meetings. When direct assessment is not possible, is it reasonable and protective to regard the estimates of exposure for the major residential and other non-occupational exposure use scenarios developed by OPP as upper bound estimates of the exposure received by infants and children from such use?

(11) In OPP's view, its aggregate exposure assessments generally do not underestimate the exposure to infants and children because the aggregate exposure is calculated by adding the high-end, probabilistic estimates of exposure to pesticides in food, to the high-end, deterministic estimates of exposure to pesticides both in water and, as a consequence of pesticide use, in residential and similar settings. Please comment on this view.

The Office of Pollution Prevention and Toxics Proposed Test Battery for the Children's Health Testing Program.

(1) Is the proposed Children's Health Testing Program battery appropriate to evaluate the potential hazards of industrial/commercial chemicals to which children may have high potential exposure? If not, what modifications are recommended?

(2) Does the SAP agree that the proposed battery should be viewed as a single tier of studies? If not, what studies in the proposed test battery are recommended as tier 2 studies and what triggers could be used to move from tier 1 to tier 2?

(3) Does the SAP/SAB have any recommendations as to the order of conduct of studies in the Children's Health Testing Program?

PANEL RECOMMENDATION

The majority of the Panel supported a weight of the evidence approach for OPP to make FQPA safety factor decisions, taking the toxicology and exposure databases into account. However, the Panel requested that the Agency provide greater clarity on the nature of the weight-of-the-evidence approach, particularly the application of the database uncertainty factor and the size of that factor. The Panel suggested that the Agency develop a standard operating procedure for acquiring and evaluating peer-reviewed studies (including human epidemiological studies) that fall outside currently required toxicology data requirements. The Agency should revisit the core

toxicology database every few years to update data requirements as needed. The Agency should also examine its testing guidelines and, where possible, combine protocols to save animal and financial resources.

Several Panel members could not determine if OPP's current approaches are sufficiently conservative and do not underestimate the risks to infants and children. In addition, OPP's proposed methodology does not adequately identify individuals that are inherently sensitive. Such individuals could be more sensitive to pesticides that lead to a number of secondary disorders (e.g., diabetic neuropathy and liver cancer). The Panel believes that OPP's decisions concerning the toxicology database uncertainty factor should not be based on the number of missing studies, but rather on the relative importance of the missing studies.

The Panel agreed with the Agency's approach of calling in data for neurotoxicity studies (acute and subchronic and developmental neurotoxicity) for conventional chemistry food-use pesticides that are known neurotoxicants, and requiring the same set of studies for proposed conventional food-use pesticides. Based on the Agency's new data requirements, it should clearly articulate when it plans to expand or reduce test requirements.

With respect to the immune system, the Panel felt that the Agency was justified in including a measure of immune function in the tier 1 testing scheme. Moreover, the Panel urges the Agency to consider a more flexible science based approach to the design and conduct of immunotoxicity studies. The Panel expressed concern that since the proposed test will be conducted in young adult animals, any developmentally-related differences in sensitivity of the immune system may be overlooked. Furthermore, the tests that are included in tier 1 do not evaluate the impact of exposure on other compartments of the immune system or on the potential for autoimmunity.

The Panel agreed with the Agency's criteria for assessing causes for increased concern for a pesticide's potential to cause developmental neurotoxicity. In addition, the criteria OPP is proposing to assess the potential for pre and/or postnatal effects are appropriate. However, the Agency should explain how each criterion would be weighted in the Agency's decision-making process.

The Panel was divided as to whether OPP's tiering process for estimating exposure to pesticides through food accurately estimates exposure to infants and children. While several Panel members concluded that the tiering system is adequately conservative to be protective for most of the population, others argued that conclusions could not be drawn without reviewing some case studies on how data from surrogates and assumptions about environmental fate are addressed in these models.

In terms of the likelihood and magnitude of pesticide contamination of drinking water and its source, the Panel questioned if the Agency's estimates represent upper bounds. The Panel concluded that if the data from modeling were intentionally biased toward upper bounds values,

then the Agency's upper bound estimates were acceptable. However, some members noted assumptions that were not conservative and questioned the degree of conservatism that could be asserted in the absence of monitoring data and further research. In addition, the Panel concluded that the Agency should not rely totally on databases to determine the potential of a pesticide to impact drinking water, due to bias in the database, but rather specific studies should be designed by the Agency.

The 10x Task Force Exposure Working Group document presented several major steps forward in the exposure assessment process that were lacking in the OPP residential exposure standard operating procedures (SOPs). In particular, there are limitations in assumptions about hand-to-mouth and object-to-mouth activities and ingestion of dust, soils, and turf in evaluating children's exposures. However, both the OPP SOPs and the Exposure Working Group document lack consistent, articulated criteria for systematic selection of assumptions. The development of the Exposure Working Group document nearly two years after the initial OPP SOPs is a reversal of the order in which these activities should have taken place. Thus, the Panel cannot answer the Agency's question as to whether OPP's estimates of residential and other non-occupational exposure to infants and children are upper bound estimates; the question implies that it is possible to judge or determine if the scenarios are reasonable and protective through empirical or semi-empirical techniques.

The Panel also discussed the EPA, Office of Pollution Prevention and Toxics Children's Health Testing Program. The Panel could not conclusively determine whether the proposed Children's Health Testing Program battery was appropriate to evaluate the potential hazards of industrial/commercial chemicals to which children may have high potential exposure. In any event, the Panel concluded that the Agency should retain the standard toxicology protocols and add the more specific developmental neurotoxicity, immunotoxicity, and neurotoxicity tests now proposed for pesticides. In addition, the Panel believes that nonpesticide (industrial/commercial) chemicals be considered in the same manner as pesticides with regard to their potential to impact the health of children.

The Panel believed it was appropriate for the proposed battery of tests to be viewed as a single tier of studies. The Agency should pursue the more standard toxicology protocols as encompassed in the proposed battery of tests. However, this position may be altered after the results of the 50 chemical surveys are evaluated. The Agency is encouraged to revisit this issue after it reviews its first group of 50 chemicals. The results on these 50 chemical studies would drive the order of conduct of studies in the Children's Health Testing Program.

DETAILED RESPONSE TO THE CHARGE

FQPA Safety Factor Issues

(1) Is a weight-of-the-evidence approach to making FQPA Safety Factor decisions appropriate, taking into consideration the toxicology and exposure databases for a

pesticide and the potential risks for the developing fetus, infant, and child as well as other populations? If not, why not? Given the scope of the evidence which OPP intends to consider, are there any other types of scientific information that OPP should consider in its FQPA Safety Factor determinations?

The majority of the Panel supported a "weight-of-the-evidence" (WOE) approach, that provides for a reasonable replication of the review and clearer understanding of the reasons behind a particular choice. The issues involved are complex, and each pesticide presents a unique combination of toxicity test results and exposure estimates. Experience has shown that the WOE approach is especially useful in such scenarios, and should be applicable to the case at hand. The Panel appreciates the concern for the Agency to appropriately interpret these data from the newly proposed toxicity tests and incorporate those findings into the risk assessment for the tolerance setting-process.

While the Panel developed consensus in the use of a WOE approach for FQPA safety factor decisions, many members were unclear as to the exact nature of the WOE approach and had reservations about its application by the Agency. This lack of clarity applies particularly to the application of the database uncertainty factor and the size of that factor, which appear to be subjective when one or more key elements of the core toxicology database are missing. There is a need for a formal mechanism for assessing available peer-reviewed literature reports of toxicity studies that fall outside currently required toxicology data sets. The Panel suggested considering a requirement for a SOP for acquiring, evaluating, and weighting peer-reviewed animal studies in the literature, and similarly for human epidemiological data concerning health effects of inadvertent chemical exposures. A transparent characterization and usage of database uncertainty factors is needed when data of these types are taken into account in the risk assessment process.

There should be a more precise definition of what is meant by WOE for the purpose of assigning a FQPA safety factor to determinations of safety. This policy is likely directed at developing reasonable assurances that a pesticide will not produce an adverse effect on health as a result of certain registered uses. The phrase "weight-of-the-evidence" is frequently used as indicating that a chemical is likely to produce an adverse effect. OPP should clearly define its use of the phrase "weight-of-the-evidence". There are likely to be other types of data that will have to be considered within the next decade that may improve the process within the next decade. Distinctions between particular types of toxicity data may become blurred. Eventually, due to the increased use of molecular biology in tier 1 screening tests, such tools are likely to better predict toxicological responses in the future.

A WOE approach to making FQPA safety factor decisions implies that expert judgment will be used to interpret uncertainty associated with toxicology, exposure and risk information. Although this sounds like a reasonable approach, the Agency should more clearly define what it means by a WOE approach, and how conditions of uncertainty would lead it to apply safety factors of different magnitudes, given different circumstances. In other words, the Panel could not conclude that the Agency has defined and adopted decision logic that will guide the choice of

additional safety factors of different magnitudes, i.e. ranging from 0-10.

If the decision logic remains as ambiguous as it currently is, it is impossible to understand the relation between available evidence, its interpretation by experts, and the choice of a specific safety factor. The Panel hopes that the Agency's reasoning will become more transparent and consistent, thereby discouraging any conclusion that the decision was influenced by other factors.

The decision logic should flow from answers to the following questions, which among others, could be used to define the ideal toxicity and exposure database:

(1) Has the Agency received and interpreted all required toxicology information for the chemical in question? (This should include developmental neurotoxicity, immunotoxicity, and effects on the endocrine system.)

(2) Does the Agency fully understand the potential for the chemical to contribute to adverse effects posed by other chemicals that are believed by its experts to act via a similar mechanism of action?

(3) For all registered uses, has the Agency received chemical release, transport, and fate data that allow it to estimate, with reasonable precision, the potential of the chemical to contaminate diverse environmental media, including food, water, air, soil, non-food plants (lawns), and indoor environments (furniture, rugs, toys, clothing etc.)?

(4) Has the Agency developed credible, probabilistic estimates of total exposure across potentially contaminated media?

(5) Has the Agency developed credible probabilistic estimates of total exposure across chemicals that are likely to act via a similar mechanism of action?

If the answer to any of these questions is "no", then the Agency faces a special presumption against relieving the 10x safety factor. This presumption is reinforced if the Agency has any reasonable basis to suggest that children, infants or fetuses are especially susceptible to adverse effects from exposure to the chemical, or group of chemicals; or if the Agency has any reasonable basis to conclude that children, infants or fetuses are more heavily exposed to the pesticides of concern than adults.

The Agency might approach the problem differently by answering the following questions:

(1) What data are necessary before it may conclude that the 10x safety factor should be relieved?

(2) What data are necessary to relieve the 10x safety factor, but not fully remove it? Since we rarely if ever will have the ideal toxicity and exposure database, the Agency will

normally face a presumption that the 10x should be retained.

(3) Could Conservative Default Assumptions Relieve the 10x safety factor?

Another question to be considered is in the absence or imperfect understanding of chemical toxicology and human exposure - could the Agency avoid applying the 10x safety factor by adopting especially conservative assumptions regarding toxicity and/or exposure, when estimating risk? The answer to this question could be yes. However at the present time, it is difficult for the Panel to understand how this might be accomplished in a consistent, transparent and scientifically defensible manner. At present, application of the 10x safety factor in the face of uncertainty is by far the simplest, most transparent approach, and one mandated by the current statute, a conclusion noted in the SAP meeting report issued following the December, 1998 meeting.

In conclusion, the Agency should define assumptions that it will adopt and apply in the absence of perfect information. A range of data availability and quality normally exists for any chemical, or group of chemicals that act via a similar mechanism. Generally, the Panel hopes that the Agency will approach the problem in the following sequence:

CJudgment regarding data sufficiency and quality.
CJudgment regarding application of conservative assumptions;
CJudgment regarding application of additional safety factor.

If the Agency concludes that data are insufficient or of poor quality, it has two options: a) apply conservative default assumptions to estimate risk or; b) do not apply conservative assumptions, and instead apply the additional safety factor. A Panel member provided the following example. The Agency is reviewing an organophosphate pesticide, currently registered for hundreds of indoor and outdoor uses. The Agency has not yet received toxicology data in the area of developmental neurotoxicity, immunotoxicity or endocrine system effects. The chemical is assumed to act via the same mechanism of action (cholinesterase inhibition) as do dozens of other registered organophosphate pesticides, however the Agency has not yet tested the toxicological effects that result from combined exposure. Environmental use and residue fate data exist for raw foods, but is limited and dated for processed foods, drinking water contamination and indoor surface contamination.

Thus the questions for the Agency based on this example are: (1) should the Agency apply conservative assumptions as it interprets the toxicity and exposure information to estimate the probable range of exposure and risk? or (2) should the Agency simply apply a default 10x safety factor?

The SAP should consider these questions when it is presented with assumptions and a logic for their application, in sufficient detail to be able to judge their conservative nature. The relations between conservative assumptions applied to uncertain information and the choice of

specific safety factors should be explored more fully in case studies. The Agency should develop these case studies, considering their most difficult regulatory situations—i.e. where they must make a choice regarding the management of a pesticide that is registered for release to diverse environmental media, and for pesticides that act via a common mechanism with other chemicals. Development of these cases would allow the SAP to more fully understand the Agency's capacity to estimate the accumulation of exposure and risk across environmental media, and across chemicals. The Agency should fully identify different sources of uncertainty in these cases. Finally, the Agency should then openly consider how its assumptions account for this uncertainty. It should then consider the relations between uncertainty, default assumptions, and the choice of safety factors.

In its FQPA safety factor determinations, the Agency is encouraged to formally revisit and review the core toxicology database every few years to ascertain if it is adequate, inadequate, or contains redundant or useless requirements. By staying current with state-of-the-art approaches, the Agency will add to the credibility of the evaluation process. Such an approach would maximize the efficiency of animal, time, and financial resources. The Agency is encouraged to examine its protocols (i.e., testing guidelines) and where possible attempt to combine protocols to save animal and financial resources. The Panel recognized that the Agency has plans in this regard but wanted to further encourage and emphasize the need for this action.

Several members believed that improved methods of neurotoxicity testing and validation of conservative assumptions regarding children's exposure would ultimately make the WOE approach a stronger tool for risk assessment purposes. However, at present, too many gaps in the available databases exist in order to be confident in decisions made under this approach. As an example, validation of the methods of exposure estimation by direct observation and measurement seems critical to the confidence in conservative assumptions. Particularly troubling are the problems with estimating hand-to-mouth, object-to-mouth activity, exposure time estimates relative to age, and soil/turf ingestion activities. These are very important routes of exposure of children at some of the most critical periods of nervous system development. Likewise troubling are the gaps in observation of critical periods of neurological development in the fetus and young, as well as a lack of understanding of the effects of endocrine disruption and neurotoxicant exposures at critical periods of early development. In addition, there is concern about the use of data derived only from the animal experimentation database. The Agency should consider data from other sources, including published peer reviewed reports in the "open" literature.

The discussion of dose-response slopes and their use in the interpretation of concern for lower doses, while statistically simple, is aimed in the right direction and suggests improved methods of analysis for non-cancer endpoints. However, the methods proposed provide only very limited evaluation of one very conservative issue, the assumption that non-cancer endpoints have thresholds. Statistical methods exist for the evaluation of the shapes of dose-response curves that can provide objective information that would be useful in evaluating this hypothesis. While one can never get a definitive answer of whether a threshold exists or not, one can estimate the

appropriate concern for the possible lack of a threshold. By applying methods which directly evaluate shape, this assumption can be strengthened (less need for the 10x factor) or weakened (suggesting possible need for the 10x factor).

NOAELs are not zero risk points; they are points at which there is greater than a 5% chance that the control and associated exposure group arise from the same distribution. There is the expectation that at the NOAEL, there is still risk. It is important to take this issue into account when evaluating the need for the 10x safety factor. For example, a NOAEL for which the possible risk (e.g. upper 95% limit) is 30% of the animals affected should have a very different bearing on the use of 10x safety factor than a NOAEL for which the possible risk is 1%. Failure to consider this issue in the evaluation could lead to substantial risks at doses considered safe, an anti-conservative risk assessment and the failure to adequately protect the public when actions are based on such an assessment.

On a related point, there is incomplete analysis of the information used to support the addition of tests to the core list of studies for tier 1. A careful analysis focusing not on NOAELs but on correlations of response patterns and magnitudes using more appropriate statistical tools would provide a clearer interpretation and provide greater scientific support for any eventual policy choice. Failure to do this analysis could leave serious gaps in the database which could lead to improper application of the 10x factor. In addition, because a NOAEL must be one of the administered doses, it is not clear that evaluations of whether certain studies lead to lower NOAELs can be properly interpreted as providing more sensitive study endpoints. A more appropriate analysis would use a standardized measure of risk, such as the ED05 or ED10 and the bounds on this estimate.

There is some confusion as to what kinds of data support the use of the 10x safety factor. For example, studies providing a strong dose-response relationship (increasing severity with increasing dose) creates greater concern for removal of the 10x safety factor. Yet, these studies generally provide the strongest information for clear identification of a low-risk exposure level and decrease the uncertainty in this estimate. Where dose-response data are inconsistent, only available for insensitive endpoints, or from studies of low statistical power, uncertainty is large and there is the possibility of unacceptable residual risk remaining after the application of the standard factors. Thus, the application of inconclusive dose-response information 10x safety factor decisions is unclear.

The uncertainty/modifying factors used in reference dose derivation are aimed at correcting for differences in sensitivity between and within species and for lack of certainty in the data. They all have fairly well defined reasons for when to use certain values for each factor. However, neither the NOAEL nor the benchmark dose represent zero risk exposure points. There seems to be the belief that these factors move from a possible risk point to a zero risk point because the factors are large. Yet this has not been demonstrated and may appear in some cases to be incorrect. The choice of the use of the 10x factor has to be addressed in light of the fact that the point-of-departure is not a zero risk point in the test species.

Another Panel member commented that the argument above could be considered purely arbitrary. One can as easily start with a benchmark dose approach instead of the NOAEL and continue to make the same argument through a whole series of 10x factors. At some point, each addition of a 10x safety factor begins to increase the uncertainty factor. It is not convincing that it increases safety. The 10x safety factor is a comfort factor, a policy factor, but it is not a factor with a lot of scientific basis. There is nothing wrong with comfort or policy factors, they just need to be identified as what they are. The same goes for the NOAEL (i.e., it is a point that arises as an accident of the experiment that was conducted). We actually do not know where the real no effect level is; that is why it is called a no observed adverse effect level.

The use of "reasonable certainty" by definition requires that the totality of the information at hand needs evaluation for making FQPA safety factor decisions. In other words, in a "weight-of-the-evidence" approach, the full range of data and evidence should be considered in making safety factor decisions.

The Panel recommended OPP should routinely obtain more specific (additional) information on a given toxicity finding. For example, if the Agency finds an endpoint of particular concern to children, it should examine the mechanistic cause or mode of action of that effect and factor the results of such findings into its safety factor determinations. In other words, the finding would provide the "trigger" for other more definitive studies.

(2) Under what circumstances, if any, do you believe that OPP's current approaches (the combination of empirical data, models, and assumptions) fail to yield risk assessments which are sufficiently conservative and do not understate the risks to infants and children?

Several members of the committee expressed concern that it is difficult to make the judgment, from the existing information as presented, that OPP's proposed procedures for FQPA risk assessments are "sufficiently conservative and do not understate the risks to infants and children." Such a judgment requires a quantitative analysis of the likely residual risks that could remain after application of OPP's procedures to chemicals that prove "positive" for developmental effects and to chemicals for which the existing testing procedures fail to detect effects. For positive chemicals, it was emphasized that the animal/human "uncertainty factor" was, for the most part, a dosimetric adjustment factor that compensates for the fact that humans tend to eliminate toxicants at a slower rate than experimental animals (with middle values tending to be approximated by the ratio of human to animal body weights to the 1/4 power—about 4 fold in the case of rats and 7 fold in the case of mice.) The generic average human/sensitive human factor of 10 fold would need to encompass somewhat more than three standard deviations in a possible lognormal distribution of human sensitivities in order to go from a 5% risk level consistent with observations of a NOAEL and a one in one hundred thousand or one in one million incidence of harm (Hattis, 1997). Recent information on the spread of human interindividual variability for mild effects in adults gives some grounds for skepticism that a tenfold factor will routinely encompass three standard deviations of a human population distribution of thresholds (Hattis, 1999).

The proposal does not identify individuals that are inherently sensitive. For example, there are a variety of multifactorial diseases for which certain chemical agents could contribute to such conditions as Parkinson's disease, essential hypertension, or non-insulin-dependent diabetes mellitus. Such individuals could well be more sensitive to pesticide agents that lead to a number of secondary disorders as apparently different as neurotoxicity (e.g., diabetic neuropathy) and cancer (e.g., liver cancer). Therefore, there are conditions in which the current process may not be sufficiently conservative because these sensitivities are not likely to be tested for in the near future with new or established chemicals.

There is additional reason for concern for populations of children and developing fetuses. In general, OPP's current approaches could fail to yield risk assessments that are sufficiently conservative if one or more of the following circumstances applies:

(1) the battery of tests in rodents and other animals used does not effectively measure a wide enough array of higher-level neurodevelopmental or other developmental functions to detect important modes of action in people.
(2) there is an insufficient allowance for human inter-individual variability to cover the diversity of human sensitivities, which in some cases may be considerably broader than the diversity of sensitivities in experimental animal populations (Hattis, 1996).
(3) there are deficiencies in estimating high end exposures for infants and children.
(4) the single-chemical risk assessment techniques fail to capture the cumulative risks from chemicals with related or possibly interacting mechanisms of toxicity.

The Panel suggests that it is important to test the degree of protection likely to be afforded by OPP's risk assessment procedures by applying them on a hypothetical basis to the observations that would be routinely produced by the required pesticide testing protocols for an array of known "positive" developmental toxicants. Such materials would include methyl mercury, lead, some specific neuroactive non-coplanar PCB congeners, and an anti-convulsive agent with known human developmental toxicity. After application of OPP's procedures for determining reference doses to the test chemicals, quantitative risks could be estimated at the reference dose (and possibly below) and the judgments could be made of the advisability of retaining the FQPA uncertainty factor for such "positive" compounds.

Some Panel members expressed particular concern that pesticides that are used in homes, daycare centers, schools, food production, and pesticides contaminating water would be likely to lead to the greatest risk in underestimating exposure from all sources and routes, as well as drive the risk relative to multiple pesticides with similar modes of action. The limited exposure assessments are well outlined in the Agency's background document. Less well acknowledged are issues of short-term exposures at critical periods of development, including those inside the uterus, as they relate to endocrine disrupting chemicals and neurotoxicants. In the absence of improved knowledge about these exposures, there should be a very conservative approach to the protection of the fetus and child.

Models and assumptions employed require validation against empirical data when such data exist and prospectively (with the planning of new studies) when they do not. Considerable uncertainties surround exposure data, particularly in infants and children, and suggest proactive and expanded acquisition of data for validation.

Scenarios can likely be developed that would involve exposures to pesticides that will predict risks greater than risks predicted by the current approaches. Panel members differed whether such scenarios are considered. OPP must specify some target percentiles of the expected population distribution of exposure for routinely evaluating whether its standard procedures provide adequate protection for relatively highly exposed people with an adequate degree of confidence.

On the other hand, some Panel members thought that the current approaches are adequately conservative and, if properly applied, should be protective of infants and children. It is the Panel's understanding that OPP will be taking into consideration potential exposures from all sources and, specifically, exposures during the entire span of human development. Further, according to OPP's interpretation of the FQPA, the default FQPA 10x safety factor must be used in the absence of reliable evidence justifying use of a different value. Use of the "risk cup" approach, which takes into account the concept of cumulative risk, i.e., the potential presence of residues of other pesticides with like mechanisms of action, adds to the conservatism of OPP's approach.

Toxicology Database Issues

(3) Please comment on OPP's proposed criteria for defining the core toxicology database and its approach to imposing a database uncertainty factor if certain key studies are missing from the database.

There was disagreement between OPP and the Agency's toxicology working group (TWG) regarding defining and implementing toxicology data requirements. The plan outlined by the TWG for implementation of data requirements, while more ambitious, seems more appropriate and protective of children. On the other hand, the apparent use of the database uncertainty factor as a replacement for the 10x safety factor required by FQPA may not always be appropriate as noted by OPP. The Panel recommends that in selecting a database uncertainty factor, the Agency consider the importance of missing studies, rather than apply a rigid default based on the number of missing studies.

The Agency appears to have a system for "weighting" the results of studies published in the open literature, but studies within the required core toxicology data-set appear to be treated equally. For example, a database uncertainty factor of 3x is applied when one key element of the dataset is missing, regardless of the identity of that key element. The Panel suggests the Agency instead weigh the importance of the missing element and make case-by-case decisions regarding the application of a particular factor.

OPP has proposed three criteria for determining which studies should be included in the core toxicology data set. The document should clearly state that data outside the three criteria presented by OPP can be used in a hazard assessment. The way it was presented almost seemed to suggest that no data would be recognized if guidelines were not provided. It was made clear in the Agency presentations that this was not the case. OPP even has guidelines that are applied to the evaluation of studies that fall outside the dictated list, and this fact should be acknowledged in the policy document.

The first criterion specifies that (1) peer-reviewed, publicly available guidelines or standardized study protocols be available and (2) there should be a scientific consensus that such a study would provide useful data for hazard assessment. Such a criterion seems reasonable, in that there is no value in requiring collection of data unless the appropriate test methodologies are well described and readily available. It is also essential that any required study should produce data that will add significantly to our understanding of the human toxicity potential of a test compound.

The second criterion contends that (1) the data from a core study should be of the type required routinely under established OPP policy and practice for either pesticide registration or reregistration and (2) the Agency has experience in evaluating such data. This criterion is primarily designed to give test sponsors both the incentive and the time to produce the necessary data. The concern here is how this criterion is to be implemented. Apparently, OPP is proposing to begin routinely requiring studies that meet its other criteria and to add them to the core requirement once some of the initially requested studies are completed and submitted to the Agency. Although this method of implementation is not ideal, it seems to be a reasonable approach in terms of practicality.

The third criterion states that there should be a scientific consensus that collection of data from such a study has actually resulted in improvement of the hazard assessments in which it has been used. However, the application of this criterion may in some cases be problematic. Scientific consensus, while desirable, may be difficult to achieve in practice due to concerns with the cost of studies. One member felt the Agency should remove the requirement of scientific consensus in applying the third criterion, and suggested the following rewording of the criterion – "whether the body of evidence supports the conclusion that information gained from the study significantly improves the understanding of the potential hazard of the pesticide to infants and children."

The Agency's approach to imposition of database uncertainty factors if studies are missing conforms to previous practice and appears reasonably conservative for application to the proposed tolerance setting process.

Finally, on a related issue, it was suggested that the Agency should consider carefully the need for requiring *in utero* cancer bioassays and investigate other possibilities (e.g., short term) to generate data on the impact of *in utero* and early-in-life exposures for both practical and scientific

reasons. The Agency needs to be aware that a rodent cancer bioassay, while complicated to conduct, asks a very simple question: "Does a given chemical possess carcinogenic activity under the conditions of the study?" While the bioassay is fairly adequate (qualitatively sensitive) for answering this question, the quantitative results are of less value. For example, it is not unusual for the incidence of a given treatment-related tumor to vary up to 2X between two studies conducted under identical conditions.

As the Agency noted, although the toxicology database is not robust, a review of the carcinogenicity of 40+ chemicals that used both the standard lifetime exposure bioassay protocol (starting at + 8 weeks) and *in utero* and /or perinatal exposure in combination with the standard protocol showed that both protocols identified potential carcinogenic activity. Therefore, if the question being answered is one of potential carcinogenic activity, the additional *in utero* exposure would not provide appreciable additional information in this regard.

There are some very complex methodological considerations that need to be considered in conducting an *in utero* carcinogenesis bioassay. If one adheres to the concept of requiring a dose that is equivalent to the "Maximum Tolerated Dose" (MTD), the MTD will probably be different for the pregnant dam than the nursing pups, post-weanling animals, and finally the 8 week old animals. This means that the exposure levels in the bioassay will potentially have to be adjusted at least four times during the study. Just as importantly, it means that four prechronic "dose-range" studies will be required before the *in utero* carcinogenesis study is conducted. These additional considerations could easily double the cost and number of animals used in the study.

Therefore, an *in utero* bioassay should be required only in special circumstances. The criteria outlined by the Agency (section D,5,c) for considering conducting such studies appears relevant for such a decision. However, no single one of these criteria would be sufficient to "trigger" such a resource intensive study. Instead a "weight-of-the-evidence" approach would be implemented.

Finally, the Panel suggested that the Agency should give thought to investigating the possibility of using a "short-term" bioassay to answer the question of the influence of *in utero* exposure on the carcinogenic potential of chemicals. If such a model were available, it would certainly be more efficient and possibly answer the question more directly and definitively than the more complex and costly *in utero* carcinogenesis bioassay protocol.

(4) After having considered the recommendations from this Panel and the Toxicology Working Group, OPP is beginning the process of calling in data for three studies (the acute and subchronic neurotoxicity studies in adult mammals and the developmental neurotoxicity study) for a subset of conventional chemistry food-use pesticides–known neurotoxicants. In addition, OPP will be proposing to require the same set of studies for all conventional chemistry food-use pesticides in the revision of the Part 158 regulations. Please comment on this two-stage approach.

The two-stage approach for expanding the newly required test methods appears to be quite logical. However, there would be substantial benefit for articulation on the basis for expanding and contracting test requirements, particularly as everything seems to be collapsing into a single tier system (i.e., no-tier). One member raised the concern that a no-tier system would not encompass all of the concerns that one would have for purposes of dose-response assessment. A tiered approach allows for consideration of processes identified in a broad screening technique but it may be of low sensitivity with regard to dose-response relationships. Such data need to have further support by a tier that is specifically aimed at establishing a dose-response relationship for endpoints most useful for making a regulatory decision. Therefore, a no-tier approach really requires a much broader effort than is proposed for confidence in the ability to perform any type of quantitative risk analysis.

At present, the need for the developmental neurotoxicity test seems to rest largely on the premise that it is at times the "most sensitive" response from a dose-response perspective. The same argument could be attributed to the endocrine system or even control of intermediary metabolism (e.g., cholesterol synthesis). Clearly, sensitivity arises from specificity in the measurements one can identify to detect adverse effects and functional endpoints with other organ systems. Nervous system evaluations come to the forefront because functional measurements are so much richer than those evaluations applied to other organ systems. The opportunity to refine the developmental neurotoxicity testing battery should not be missed. While the endpoints in the current battery assess the integrative functioning of the sensory, motor, and cognitive systems with supportive neuromorphology measurements, the limited exposure via the mother/dam may not provide adequate or accurate levels of exposure to the offspring to assess neurotoxicity. Aspects of the dosing paradigm to consider are the extension of exposure to postnatal day 21 (consistent with the OECD guidelines), direct administration of the compound to the offspring after birth, and shorter intervals of exposure, including acute exposure during development.

The Agency should consider the practical aspects of modifying protocols to provide multiple endpoints with any one-study protocol. The Agency is beginning the process of calling in data for the developmental neurotoxicity study for a subset of conventional chemistry food-use pesticides known for neurotoxicity. There is a certain logic in using known neurotoxic pesticides as the initial test cases from which to gain knowledge and experience in the evaluation of data from the newly required neurotoxicity studies. That is the case because there should be a greater likelihood of at least some degree of neurotoxic effects observed in tests of this subset of pesticides. However, there is also the likelihood of bias from this data set of known neurotoxicants. Alternatively, the Agency should consider that selecting a few pesticides from the universe of those that do NOT act by neurotoxicity mechanisms could be instructive for comparison with representative samples of the neurotoxicants (e.g. organophosphates, carbamates and synthetic pyrethroids) in the developmental neurotoxicity studies. This would allow the Agency to more accurately assess the sensitivity gained with the developmental neurotoxicity data.

The Panel felt that the Agency was justified in including the evaluation of the immune

system as part of tier 1. Guidelines for immunotoxicity testing already exist with regard to chemicals (OPPTS 870.780) as well as the biochemical pest control agents (OPPTS 880 series). It is particularly significant that a functional test for immunity be included in this data set. A test that challenges the immune system to respond (such as the antibody response to sheep red blood cells) is appropriate. The assay that utilizes this antigen has undergone extensive validation. Furthermore, a considerable database exists with regard to pesticide exposure on this response in experimental animals.

It is recommended that the Agency consider a flexible science-based approach to the design and conduct of immunotoxicology studies by carefully considering the results from the other tests proposed in tier 1 that identify other potential target organs and consideration of potential for recovery or transient effects. It is cautioned that currently, predictive animal models for autoimmunity are not well developed and the paucity of biological information on the developing immune system represent limitations of the identification and inclusion of such endpoints into a testing protocol. The Agency should continue its efforts to develop and validate protocols that are designed to evaluate the potential for chemically-induced developmental immunotoxicity. The goal should be the creation of a carefully designed developmental toxicity study that incorporates the evaluation of functional immunity.

(5) The OPP Policy Guidance indicates that one of the critical issues is whether or not to apply an FQPA Safety Factor pending receipt of newly-required studies. In the absence of the results from any of the studies to be required through data call-in notices (i.e., the acute and subchronic neurotoxicity studies in adult mammals and the developmental neurotoxicity study), what information from existing studies on a specific chemical would increase or decrease your concerns about the potential for pre- and post-natal hazard, in general, and for neurotoxicity and developmental neurotoxicity, in particular? Which, if any of the seven criteria discussed in section V.A.1.a., footnote 4, and associated text of the OPP Guidance, are appropriate for judging whether there is increased concern about the potential for a pesticide to cause developmental neurotoxicity?

All of the criteria mentioned in Section V.A.1.a, footnote 4 and associated text, seem reasonable, although it is not clear how each criterion should be weighted for decision-making. The list of seven criteria proposed by OPP appears to be useful under at least some circumstances and covers the types of information that might be derived from existing studies that would affect the degree of concern about a compound's potential for developmental neurotoxicity. The two additional indicators in the footnote are likely to be more sensitive and perhaps more specific for identifying agents of concern. The exception noted as “unless other information…” seems very vague and opens the possibility of great misinterpretation. It should be eliminated or made much more specific and restrictive. The criterion based on potential endocrine disrupting effects should be invoked in a relatively liberal manner until there is more information available about the characterization of and specific effects of endocrine disrupting effects. In addition, it is unlikely that significant data on learning and memory processes are available for existing compounds. Developmental neurotoxicology testing should be required now, the data should be reviewed, and

the need to require the testing revisited after a defined period of time to assess it's impact on improving characterization of risk.

Additional information from existing studies that would increase concern levels for potential effects on the immune system include evidence for increased incidences of infection or of allergic responses, as well as evidence of tumorigenicity. The latter observations would most likely be seen in chronic or lifetime exposure studies. In addition, several other criteria are suggested that might reasonably be considered to increase suspicion for developmental effects in general and neurodevelopmental effects, in particular:
(1) inhibition of cell division (e.g., colchicine).
(2) specific toxicity/lethality for dividing cells (e.g., ionizing radiation).
(3) changes in neuronal migration (e.g., methyl mercury).
(4) neuroreceptor/neurotransmitter agonism or antagonism.
(5) molecular resemblance of parent compounds or predictable metabolites to known neurotoxins (e.g., gamma diketones such as 2,5-hexanedione; certain nitriles/cyanide compounds; some metals and organometallic compounds, such as alkylmercury, lead, manganese, cholinesterase inhibitors).
(6) high lipophilicity conducive to concentration in lipid bilayers important for neural functioning (e.g., PCBs).
(7) identification of decreased biological factors in the adult that could present a problem in the developing organism (e.g., decreased cholesterol with carbon disulfide could be significant for the developing nervous system due to its high demand for cholesterol).
(8) mode of action on the target species and its relationship to the human system, whether directly or via an associated mechanism or human homologue.
(9) mutagenicity, clastogenicity, or carcinogenic responses may increase concerns as well because of the implications that these effects have for low dose extrapolation.
(10) clear positive results from the two-generation reproduction studies and prenatal developmental toxicity studies in the absence of maternal toxicity would increase concern about pre- and post-natal hazards.

(6) Please comment on whether you expect that the NOAELs that are identified in the developmental neurotoxicity studies would, for a substantial number of chemicals, be lower than those NOAELs identified in the suite of studies historically required and used for age-related comparisons and Reference Dose derivation (e.g., prenatal developmental toxicity or multigeneration reproduction study, subchronic and chronic studies, etc.). Please explain the basis of your opinion.

The Panel could not develop consensus whether NOAELs from developmental neurotoxicity studies would be lower than from historically required studies. One member agreed strongly that the NOAELs or more appropriate bench mark doses identified by the developmental neurotoxicity studies will be lower than those detected by the present tests for a substantial number of pesticides. This prediction was based on the fact that the effects of many teratogens (e.g., psychoactive compounds, anti-seizure medications, anticarcinogens, metals, radiation, retinoids, folate levels, etc.) are already known to be detected at lower doses with these tests than

with the ones presently required. Another member did not accept the notion that the number of chemicals with effects occurring at lower doses would be large, but agreed that those identified would represent an important group. In addition, the analyses already presented by the Agency (Makris et al, 1998) indicate that NOAELs identified by the use of developmental neurotoxicity testing are often not likely to be lower than those characterized by prior testing methods. The Panel is aware that only one of 12 chemicals showed developmental neurotoxicity effects at lower doses than were observed with the prior standard testing protocol. However the Panel expressed caution that the results from testing the 12 pesticides could not be applied to a broader set of pesticides.

One member questioned the wisdom of moving tests of central nervous system function into tier 1, with no plans for testing the functions of other organ systems. Another question regarding the proposed battery was whether it is intended as a screen or as research. It was stressed that the results of the developmental toxicity study must be usable for risk assessment.

Several members supported the idea that the Agency needs to improve and refine the proposed battery. Because new factors in development are being discovered at a rapid rate, the Agency needs to be flexible, and the pace of development, validation, acceptance, and implementation of new protocols needs to be increased.

(7) OPP is proposing to adopt the framework and its criteria/factors for assessing the degree of concern about the potential for pre- and post-natal effects as recommended by the 10x Task Force Toxicology Working Group. Please comment on the appropriateness of the proposed criteria/factors for use in this assessment process and on OPP's proposed approach for accommodating its concerns in the Reference Dose derivation and FQPA Safety Factor decision processes, in the near term and in the longer term. What scientific considerations relate to the addition of a safety factor, where the hazard to infants and children is well characterized and the data show that infants and/or children are more susceptible than adults?

This is a difficult question unless there is some allowance for what might be loosely construed as a severity-of-effect determination. To scientifically determine such weights, there needs to be some relatively well considered process for establishing these factors for different outcomes. While the endpoints could differ, a scale would need to be developed for effects resulting from pre-and postnatal exposure that is essentially the same as that of the adult.

In general, the criteria OPP is proposing to use in assessing the degree of concern about the potential for pre- or postnatal effects (as shown in Table 4 of "The Office of Pesticide Programs' Policy on Determination of the Appropriate FQPA Safety Factor(s) for Use in the Tolerance-Setting Process") are appropriate for the intended purpose. There may be some indications of greater variability in children's responses to pharmaceutical agents, but the Panel is not certain how relevant this relative variability is to the distribution of sensitivities that are produced by genetic variation. The Panel believes that the latter is much more important than the

former when discussing environmentally-induced disease. The question is not whether some additional safety factor needs to be applied for children, but whether the uncertainty factors adequately account for variability in the general population. Although this question has never been adequately evaluated, it is central to issues in environmental health. These additions would avoid the necessity of forcing an either/or decision when, as is sometimes the case, the available data are difficult to interpret and thus not clearly of either higher or lower concern.

Similarity between animals and humans increases concern, while dissimilarity decreases concern regarding toxicity seen in animal models. Thus, lack of adequate data on toxicokinetics or mechanism of action would add some degree of uncertainty and should fall between the "higher" and "lower" extremes. Not only would the possibility remain that if such data were obtained that would show similarity to the human condition, but also treating the lack of such information as being of no consequence would provide an incentive not to study comparative toxicokinetics or mechanisms of toxicity. The Agency should note though that in cases where there are clear toxicokinetic differences between humans and the experimental animal, the agent may not have been adequately tested in a relevant species and may be indicative of important missing information.

OPP's approach could be made more readily understandable by inclusion of one or more flow charts in the Agency's background document. These should highlight decision points, the kinds of inputs considered at each such point, and the possible alternative outcomes. Separate charts could, for example, illustrate the past, current, and proposed approaches, and should illustrate the entire process, including the incorporation of exposure data, leading to the final regulatory outcome.

Exposure Issues

(8) Subject to the qualifications expressed in the OPP Policy document and the report from the 10x Task Force Exposure Working Group, OPP believes that each of the tiers for estimating exposure to pesticides through food, in almost all instances, will not underestimate exposure to infants and children. Please comment on this conclusion, as it applies to each of the tiers.

The Panel was divided in response to this question. While several Panel members concluded that the tiered system is conservative in the sense that it would be protective for most of the population, others argued it is hard to draw conclusions without seeing some case studies on how data from surrogates and assumptions about environmental fate are built into these conservative models. While it does not make sense to demand absolute knowledge on exposure, one must be made comfortable with the process before opinions can be rendered.

Although the SAP recognizes that the Agency knows more about food based exposure to pesticides compared to inhalation and dermal uptake, the Panel does not agree that current methods “will not underestimate exposure to infants and children”, as will be described below.

Many factors govern the quality of estimates of pesticide exposure in foods. These include:

Age of data: The Agency is gradually acquiring more recent food intake data, however it still relies on data that are several decades old when it estimates dietary exposure.

Sample size of age classes: Age groupings that may experience high exposure during periods of high susceptibility are commonly under represented in food intake surveys. This limits the Agency's capacity to estimate exposure and risk for groups, including infants and children less that 5 years in age.

Demographic stratification: Food intake surveys are not stratified within age classes relevant for identifying those at special risk. In other words, the Agency does not know if exposure among infants varies significantly by income level, ethnicity, region of the country, and season. All of these factors are relevant to the choice of a safety factor.

Accuracy of recipe files: Foods reported eaten by those surveyed are broken into more fundamental foods that are regulated for pesticide contamination. This is accomplished by a recipe file that breaks pizza pie, for example, into components such as wheat, cheese, tomato paste, etc. Individual recipes vary considerably among products, and change constantly as new foods are introduced. The Agency has not estimated the magnitude of the effect that imprecision in the recipe file may have on pesticide exposure via food ingestion.

Use of percent crop treated data: These data are not available for public review, and they are commonly employed to adjust chronic exposure and risk estimates. If 20% of the national apple market is treated by pesticide X, and if its distribution is primary local (i.e. not uniformly distributed nationally), then reduction of "national exposure and risk estimates" by 80% significantly underestimates exposure for 20% of the population. Most cases are far more complex, as chemical use patterns vary, and food product markets vary. Thus, application of this "exposure reduction factor" is difficult to justify, and will commonly underestimate risk for subpopulations that may include children.

Changes in Marketing and Processing: New marketing and packaging practices can change food intake patterns quickly, especially for children (e.g. juice boxes leading to increased juice consumption; increases in blends of fruit concentrates in juices and blends of vegetable oils; boxes of clementines recently introduced from European markets).

Water: Water contamination from pesticides was recently surveyed by the U.S. Geological Survey, and found to be far more extensive than previously recognized. Contaminated water clearly has the potential to increase exposure via drinking water, but also via the addition of water to food concentrates, dried foods, grains, etc. The Agency has not demonstrated the potential contribution of contaminated water to food-based exposure, especially the regional variation that might be anticipated. The SAP recognizes the importance of water as the most consumed food in

the human diet, and it recognizes that contaminated water contributes to human exposure via ingestion, inhalation and dermal uptake.

Pesticide Residues: A significant source of uncertainty in dietary exposure estimates grows from the sampling design of federal surveys of residues in imported and domestically produced foods. These surveys provide a limited view of residues in the food supply for several reasons: 1) Sample sizes for specific pesticide-food samples are normally extremely small relative to the volume of food in the marketplace; 2) Processed foods receive limited attention from FDA and USDA; 3) pesticides that require individualized tests are more rarely sampled than pesticides detectable via multi-residue screens; 3) An increasing proportion of the US food supply is derived from foreign sources, constantly expanding the "universe" of pesticides that might be on imported foods; and 4) Blending portions of crops selected from different pieces of fruit, or from different crates or shipments will systematically underestimate pesticide residue levels.

The Agency should strive to develop data that permit it to estimate both acute and chronic exposure for individuals. This should best be done by aggregating exposure across the foods that they have reporting eating for individual days. Consecutive 3 day sampling strategies will not sufficiently capture intra-individual variation across time for the purposes of chronic exposure estimation, especially if sample sizes are small for the study of relevant age groups (infants, children 1-2, etc.). The most desirable outcome would be to estimate both acute and chronic exposure as probability distributions for relevant age groups

The absence of clear standards to judge quality of data make it difficult for the Agency or the SAP to judge the magnitude of uncertainty that exists in estimates of exposure from food and other sources. The SAP encourages the Agency to focus limited resources available for food ingestion research to better understand key contributing sources to exposure *in utero* and during the first 5 years of life. This could be accomplished by looking at the cluster of chemicals expected to be found on foods most consumed by children and pregnant women.

Within different exposure scenarios involving food, there will always be some probability of highly contaminated foodstuffs getting through the screening process for contaminants. For example, how can individuals be protected from spillage of highly concentrated pesticide in storage. Perhaps only one apple was contaminated. Therefore, no screening system and no affordable analytical scheme are now available for what is essentially an accidental poisoning. As a result, it does not make sense to develop a national regulatory program around such extremes. On the other hand, it is not practical to think that meaningful empirical data will exist on exposure before a chemical is introduced into the market. Consequently, the program must identify the minimum size of the group that might be impacted by usual consumption of foodstuffs at some maximum level of probable contamination. It seems reasonable to base these projections on pesticides with similar chemical and physical properties and usage patterns.

(9) OPP is developing a tiered approach to assessing the likelihood and magnitude of contamination of drinking water and its sources by pesticides. The Panel has been asked to

comment on aspects of this activity at previous SAP meetings. As an interim approach, when direct assessment is not possible, is it reasonable and protective to regard the estimates generated by OPP's current methodology as upper-bound pesticide concentrations for surface and ground water and to assume that this concentration will be found in drinking water?

Prior reviews considered the approach as sufficiently conservative. However, the models appear to be most useful for identifying those pesticides that are unlikely to reach water in appreciable concentrations. Departure from the upper-bound estimates by virtue of examining exposure in current databases must be done with caution.

The question posed is that these estimates would be upper-bound estimates for surface and ground water. It certainly starts out that way if the original modeling is done by deliberately biasing the analysis toward "upper bounds". However, if OPP depends on measurements in the surveys identified for refinement of these estimates from monitoring data, it is no longer clear what the upper bound is. Some of the databases referred to are far from random samplings. The bias in these data systems has long been recognized. Some have been biased toward picking up positives, especially data in groundwater. Some of these data systems do not even identify whether the sample came from the drinking water or the source water. In a significant number of cases, it has not been possible to identify the source of drinking water because many cities depend upon several sources. Bias has also been introduced because negatives may simply come from areas where a pesticide was not used. Information on drinking water samples seldom identifies the treatment processes. Data that have been developed for compliance under the Safe Drinking Water Act are recorded centrally only if a maximum contaminant level (MCL) has been exceeded. without adequate definition of quality even today. A valid test of a pesticide's impact on a water supply based upon actual data requires first the opportunity (i.e., local use) and then the appropriate properties to be mobilized. Further, it is absolutely necessary to understand the treatment processes used in sampled drinking water systems before the results can be generalized -- even to other systems that use the same source water.

Consequently, it is probably not appropriate to rely on databases to determine the potential of a pesticide to impact drinking water. It would seem that the only way that this question can be answered with sufficient rigor is to design studies to specifically evaluate this question. Dependence upon existing or even future databases that may be more representative may, in fact, not represent the use patterns associated with a particular pesticide, even though the database could be representative for the country as a whole.

One final issue is that significant exposures to pesticides are likely to be episodic. Large systems are unlikely to end up with significant exposures for many reasons. Better water treatment and large volumes increase the likelihood of dilution and other considerations. Small systems could be exposed to a spill located close to source, encounter storm events that might introduce particulate matter into the water, have high levels of local irrigation, etc., that all increase the vulnerability of drinking water to pesticides.

Another extreme exposure that needs to be considered may be a migrant worker's child swimming in and drinking water from irrigation ditches, etc. It is not clear that these scenarios play out very strongly in deriving upper bound exposures of pesticides in drinking water.

(10) OPP is developing approaches to assess the likelihood and magnitude of exposure to pesticides in residential and other non-occupational-use scenarios. The Panel has been asked to comment on aspects of this activity at previous meetings. When direct assessment is not possible, is it reasonable and protective to regard the estimates of exposure for the major residential and other non-occupational exposure use scenarios developed by OPP as upper bound estimates of the exposure received by infants and children from such use?

The 10x Task Force Exposure Working Group should be commended for the background document they have developed. It advocates a number of major steps forward in the exposure assessment process that overcome major shortcomings in OPP residential exposure standard operating procedures (SOPs), including the incorporation of probabilistic approaches, the recognition of narrowly defined age groups relevant to specific exposure-related behaviors (i.e., prenatal, crawlers, young toddlers, etc.), movement of pesticides across media (e.g., deposition on non-target surfaces), and recognition of the importance of receptor-based (as opposed to source-based) exposure assessments that examine important exposure issues from the perspective of how and where children spend time.

The Panel urges OPP to fully integrate the above steps into the exposure assessment process for non-dietary exposures. Indeed the production of this document, nearly 2 years after the initial residential SOP protocols is a reversal of the order in which these activities needed to have taken place, meaning that it is difficult to answer the question put forward to the Panel because the question implies that it is possible to judge or determine through empirical or semi-empirical techniques if the scenarios as articulated in the document are reasonable and protective.

Whether scenario-based residential and non-occupational exposure assessments are sufficiently conservative so as to not underestimate exposures hinge on several issues:

(1) whether the scenarios chosen are exhaustive, i.e., have included every potential possible exposure scenario and have not overlooked cross media transfer.
(2) whether measurement and assessment data and exposure factors are accurately characterized.
(3) whether exposure factors based on data and default assumptions have been chosen in a consistent manner and reflect within individual variability in behaviors so that assessors know whether or not contact rates and durations are truly upper bound.
(4) the timing of exposures relative to one another, given that many pesticide applications take place on a seasonal basis. It is possible that exposures by more than one scenario (e.g., turf applications and wading pool exposures) can take place within a day or days of each other.

Multiple concerns have been raised by the SAP and other groups regarding the inadequacy

of the residential SOPs, particularly weaknesses in assumptions about hand-to-mouth and object--to-mouth activities and ingestion of dust, soils, and turf. The SOPs and exposure assessment process as described in the Exposure Working Group paper have a number of shortcomings related to the lack of consistent, articulated criteria for systematic selection of assumptions. The document notes that "conservative scenario mixes median and upper-bound exposure factors" but this is often applied in a haphazard fashion or ignores median values for key data sources where they exist. The goal should be to use the median values of well articulated exposure distributions (body weights or surface areas, for example) and choosing conservative but defensible upper-bound estimates where chemical-specific data do not exist (e.g., 100% inhalation absorption). Examples of scenarios where this is true include: 1) the use of the 1.56 hand contact rates per hour when median values of two well conducted studies show that the true median is closer to 10; 2) use of a 15-kg body weight for 1-6 year olds in several scenarios. In other cases, indefensibly conservative assumptions are used, such as 350 cm^2 for hand surface area in hand-to-mouth ingestion scenarios, a number that includes both sides of the hand and the surfaced area in between the fingers, or the handful-of-grass consumption assumption, which appears to have had little thought put in to it.

(11) In OPP's view, its aggregate exposure assessments generally do not underestimate the exposure to infants and children because the aggregate exposure is calculated by adding the high-end, probabilistic estimates of exposure to pesticides in food to the high-end, deterministic estimates of exposure to pesticides, both in water and as a consequence of pesticide use in residential and similar settings. Please comment on this view.

The major issue here is how combining data of varying quality (i.e., food, water, non-dietary) with widely different confidence intervals affects the end result. Deterministic approaches are not necessarily always more conservative than assessments that use distributional approaches, especially when the data sets for concentration, contact rates, and duration are robust. This is not a reasonable view in light of the severe defects in the assessment of non-dietary exposures of the fetus, infants, and children. As a result, there is no confidence in the assessment of aggregate exposure.

The Office of Pollution Prevention and Toxic Substances Proposed Test Battery for the Children's Health Testing Program.

(1) Is the proposed battery for the Children's Health Testing Program appropriate to evaluate the potential hazards of industrial/commercial chemicals to which children may have high potential exposure? If not, what modifications are recommended?

The Panel was divided on its review of the proposed battery of tests. On the one hand, it was recognized that it would be ideal to have the most sensitive tests possible to detect potential hazards to children. It is recognized that the standard toxicity tests are especially weak in their sensitivity to developmental effects. However, there is always a tradeoff between breadth of assessment and specificity. At this time, the Panel believed that it was prudent to retain the

standard toxicology protocols for their breadth and add the more specific developmental neurotoxicity, immunotoxicity, and neurotoxicity tests now proposed for pesticides -- these address areas known to be missed by the old protocols and tap functions known to be subject to injury in developing humans.

Several members emphasized that future protocols should include testing end points during development and testing of animals exposed to acute and intermediate dosing. These additional requirements are crucial to evaluations of developmental toxicity.

The Panel believed that nonpesticide (industrial/commercial) chemicals should be viewed in the same light as pesticides with regard to their potential to impact the health of children. In other words, the toxic responses in animals would be expected to be the same for an industrial chemical as a pesticide of similar chemical structure/activity. That being the case, it would be prudent for the Agency to require the same or similar types of toxicity data on chemicals of industrial/commercial use as pesticides.

There is one essential difference between industrial/commercial chemicals and pesticides; the universe of industrial/commercial chemicals is much larger than for pesticides. Therefore, priority setting for industrial/commercial chemicals will be a preeminent consideration. One member suggested that the primary criteria in choosing chemicals to be tested should include: 1) those chemicals where exposure to children would be expected to be high compared to adults, 2) chemicals where children are uniquely exposed (i.e., large numbers of exposed children) and; 3) chemicals where there is concern about unique sensitivity to the toxic effects of the chemical. The Panel member believed that all three criteria should carry more weight than production volume, although this could also be considered in their selection, and that final consideration should be given to evaluating chemicals for which there is a rich database, at least initially. The Agency noted that data on developmental neurotoxicity, immunotoxicity, and some other measures are available for very few of the items on the list. Metabolism data are often minimal. That is, there are essentially no compounds for which the database is good in regard to children's health. EPA's goal should be to get a consistent set of data on 50-60 chemicals where there is reason for special concern, then re-evaluate the value of the tests.

(2) Does the SAP agree that the proposed battery should be viewed as a single tier of studies? If not, what studies in the proposed test battery are recommended as tier 2 studies and what triggers could be used to move from tier 1 to tier 2?

The Panel believed that it was appropriate for the proposed battery of tests to be viewed as a single tier of studies, at least initially. However, the Panel was divided on the "mix" of the proposed battery of tests. On the one hand, it was recognized that it would be ideal to have the most sensitive test possible to detect a potential hazard to children. It is recognized that the proposal includes adult tests that may be inadequate to determine children's health. By their nature, "sensitive" tests are fairly specific with regard to their endpoint and, therefore, may preclude finding other outcomes of exposure to the chemical. In contrast, more general types of

studies (e.g., acute, 90-day, etc.) have the ability to evaluate large numbers of endpoints, but may miss a subtle effect.

At this time, the Panel believes that it was prudent to pursue the more standard toxicology protocols as encompassed in the proposed battery of tests. However, this position may change after the results of the 50 chemical surveys are evaluated. The Agency is encouraged to revisit this question after it reviews its first group of 50 chemicals for which there is information readily available on the proposed battery of tests. It is apparent that few of the 50 chemicals will have data on all of the tests in the battery. However, the Panel concluded that, based on Agency input, enough data would be available on enough of the chemicals to construct a matrix that would give insight into the value of the proposed batter for predicting risk to children.

After reviewing the results of the "matrix" evaluation, the Agency might find a need to require other studies on a given chemical to evaluate the potential hazard to children. Additionally, if the data suggest that a given chemical is a potential toxin, then the Agency might want to require specific tests to define the sensitivity (dose response) and characterize further that specific endpoint. The Panel is of the strong opinion that this process needs to be a "Work in Progress" with timely critical reviews. In this respect, it should be viewed as an evolutionary process.

The Panel suggested that the Agency should take this opportunity to develop testing protocols to evaluate functional alterations following developmental exposure. In addition, the Agency needs to give thought to the timing and length of gestational exposure, e.g., intermittent vs. acute, for chemicals that have the potential to produce neurotoxic effects.

(3) Does the SAP/SAB have any recommendations as to the order of conduct of studies in the Children's Health Testing Program?

The consensus of the Panel was that it is premature to "order" the conduct of the studies. At this point, there simply is not enough information to provide credible advice. The Panel believed that the results of the 50- chemical study would logically drive the ordering of studies in the future. In the meantime, the Panel though that there would be benefit to the "staging" of studies for chemicals for which data are lacking. For example, studies that require a shorter period to conduct would be "first-in-line". However, the Agency should maintain flexibility to "order" or "reorder" studies as required by the issues and findings at hand at that time.

REFERENCES

Hattis, D. "Variability in Susceptibility -- How Big, How Often, For What Responses to What Agents?" Environmental Toxicology and Pharmacology, Vol. 4, pp. 195-208, 1997.

Hattis, D.; Banati, P., and Goble, R. "Distributions of Individual Susceptibility Among Humans for Toxic Effects-For What Fraction of Which Kinds of Chemicals and Effects Does the Traditional 10-Fold Factor Provide How Much Protection?" Presented at the International Workshop, Uncertainty in the Risk Assessment of Environmental and Occupational Hazards, Bologna, Italy September 25-26, 1998, Annals of the New York Academy of Sciences, 1999, in press.

Hattis, D. "The Challenge of Mechanism-Based Modeling in Risk Assessment For Neurobehavioral Endpoints." Environmental Health Perspectives, Vol 104, Suppl. 2, pp. 318-390, April 1996.

Makris, S.; Raffaele, K. Sette, W. and Seed, J. "A Retrospective Analysis of Twelve Developmental Neurotoxicity Studies Submitted to the USEPA Office of Prevention, Pesticides and Toxic Substances (OPPTS)". USEPA. November 12, 1998.

APPENDIX

The Panel proposed the following specific additions to the "Moderate" column under "Degree of Concern" in Table 4 of the Agency's background document:

(1) In the "Human data on pre- and postnatal toxicity" row, insert "Equivocal or suggestive effects that may be related to exposure."

(2) In the first row of the "Dose response nature of the experimental animal data" section, insert "Incidence or intensity of response equivocal but suggestive of a dose-response."

(3) In the first row of the "Relevance of the experimental animal data to humans" section, insert "Comparative toxicokinetic data inadequate or unavailable."

(4) In the last row of the Moderate column, insert "Mechanism of action uncertain or unknown."

(5) Incorporation of Part VI of the "Standard Operating Procedures for the Health Effects Division FQPA Safety Factor Committee" into the OPP policy document, perhaps as an addendum, would also help to clarify the proposed methodology.

SAP Report No. 99-03B, May 26, 1999

REPORT:
FIFRA Scientific Advisory Panel Meeting, May 26, 1999, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment

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

Christopher Portier, Ph.D
Chair
FIFRA/Scientific Advisory Panel
Date:_____

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
May 26, 1999

SESSION II: Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment

PARTICIPANTS

Chair
Christopher Portier, Ph.D., National Institute of Environmental Health Sciences, Research Triangle Park, NC

FIFRA Scientific Advisory Panel Members
Ronald J. Kendall, Ph.D, Professor and Director, The Institute of Environmental and Human Health, Texas Tech University/Texas Tech University Health Sciences Center, Lubbock, TX
Herb Needleman, M.D. , Professor of Psychiatry and Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA

FQPA Science Review Board Members
Chris Frey, Ph.D., Assistant Professor, North Carolina State University, Raleigh, NC
Dale Hattis, Ph.D., Professor, Clark University, Worcester, MA
John Wargo, Ph.D., Associate Professor, Yale University, New Haven, CN
Mark Whalon, Ph.D., Professor, Michigan State University, East Lansing, MI
Lauren Zeise, Ph.D., California EPA, Berkeley, CA

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

PUBLIC COMMENTERS

Oral statements were received from:
Leslie Bray, Ph.D., (American Crop Protection Association)
Robert Sielken, Ph.D. (American Crop Protection Association)
Leila Barrajm, Ph.D. (Novigen Sciences, Inc.)

Written statements were received from:
American Crop Protection Association

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency regarding statistical methods for use of composite data in acute dietary exposure assessment. Advance public notice of the meeting was published in the *Federal Register* on May 5, 1999. The review was conducted in an open Panel meeting held in Arlington, Virginia, on May 26, 1999. The meeting was chaired by Christopher Portier, Ph.D. of the National Institute of Environmental Health Sciences. Mr. Larry Dorsey, SAP Executive Secretary, served as the Designated Federal Official.

EPA has identified a reliable, statistical methodology for applying existing information from the U.S. Department of Agriculture's (USDA) Pesticide Data Program (PDP) report to acute dietary risk assessments. This statistical methodology extrapolates pesticide residue data from composite samples of fruits and vegetables into single units of fruits and vegetables. Given the composite sample mean ($\bar{X}$), the composite sample variance (S^2), and the number of units in each composite sample, the methodology estimates the mean and variance (: and F^2) of the universal distribution of pesticide residues present on single units of fruits and vegetables. With these parameters and the assumption of log-normality, values of pesticide residues on individual units are generated and then applied to a Monte Carlo probabilistic calculation of dietary risk assessment.

Mr. Dave Miller (EPA/Office of Pesticide Programs) presented a brief introduction that explained the background and needs for an extrapolation methodology. Dr. Hans Allender, (EPA/Office of Pesticide Programs) provided a detailed summary of the methodology and described the technical aspects of the statistics involved. Dr. Linda Abbott (USDA) introduced several technical points as a prelude to questions be presented to the Panel.

CHARGE

The specific issues to be addressed by the Panel are keyed to the background document entitled *Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment* and are presented below:

(1) Measurement of many natural processes may be described by typical statistical distributions, e.g., normal, lognormal, etc. In previous data-fit studies, data on concentrations of residues on fruits and vegetables have been fitted to a lognormal distribution. The lognormality of residues has been established as a fundamental assumption in the decomposition procedure. Please comment on the assumption of lognormality.

(2) The application of OPP's decomposition methodology calls for at least 30 "detects". This is done to assure that there is enough representation in the sample and that the extrapolation will cover the width of the distribution of single units. Although 30 detects is a practical rule for the

application of the procedure, please comment on the consideration of other numbers as a practical rule of application.

(3) The standard deviation within a composite cannot be greater than the standard deviation of the population of individual residues. Are there any circumstances where the standard deviation within a composite can be greater than the standard deviation of the population of individual residues?

(4) OPP acknowledges that the collection of composite samples in the USDA Pesticide Data Program (PDP) protocol is not purely random; therefore, the decomposition procedure will produce an overestimation of the standard deviation of the lognormal distribution of residues on fruits and vegetables. Moreover, the overestimation of the standard deviation is accentuated to the degree that the collection of composition samples departs from pure randomness. The consequence of overestimating the standard deviation is that the high end of the estimates of residues in single units may exceed what occurs in reality. What criteria (if any) should be used to establish an upper-bound on the amount of residue projected in a single unit to address the potential for overestimation of the standard deviation?

(5) OPP's methodology is sensitive to the number (N) of single units/servings of a commodity estimated to be in a composite sample. Please comment on how to estimate the number of single units/servings per composite sample. (Consider how to handle fruits for which a single unit is typically only a part of a unit of a commodity e.g. a melon), or many different units [e.g grapes], even though the single unit is smaller than the typical composite sample).

(6) The decompositing procedure estimates the number of units in a PDP composite by dividing the weight of the composite by an average weight of an individual unit. The number of individual units in a composite will vary, depending upon the weight of each composite unit. Will differences in the number of individual units in a composite introduce substantial uncertainty?

PANEL RECOMMENDATION

A fundamental principle of the Agency's decomposition procedure is the assumption of lognormality. The consensus of the Panel was that even though lognormality is a reasonable beginning for the distribution of underlying single-sample residues, lognormality would not generally be expected for the distribution of residues found in composites. Even if the distribution of underlying single-sample residues was perfectly lognormal, the residue found in each composite is effectively a weighted arithmetic mean of some number of single samples; a weighted arithmetic mean of lognormal samples is not expected to be lognormal itself.

OPP's decomposition approach indicated that at least 30 detects are necessary for application of the methodology. The Panel concluded that a usable analysis could be possible based on data sets with fewer than 30 detects; a minimum number should not be required for application of the decomposition methodology.

The Panel agreed with the Agency that the standard deviation of residues on individual samples making up a composite would not be expected to be greater than the standard deviation of residues in a national sample of individual single servings.

Differences in composite units could introduce uncertainty into the analysis. Consideration of this uncertainty can be addressed by collecting data on how the number of units per composite varies among composites for specific commodities, followed by numerical experiments simulating effects on the calculations. In addition, the numerical procedure to estimate the number of single units/serving per composite sample should correspond to the sampling procedure used for construction of the composites.

The Panel was encouraged by the data provided by the public commenter, Dr. Robert Sielken. Even though the Panel did not have the opportunity to critically review the information, the Panel recommends that Dr. Sielken publish the procedure and examples of its implementation in a peer-reviewed journal. Following this, the Agency should actively explore the feasibility of using it or adapting it for the exposure estimation problems that were the focus of the session.

DETAILED RESPONSE TO THE CHARGE

(1) Measurement of many natural processes may be described by typical statistical distributions, e.g., normal, lognormal, etc. In previous data-fit studies, data on concentration of residues on fruits and vegetables have been fitted to a lognormal distribution. The lognormality of residues has been established as a fundamental assumption in the decomposition procedure. Please comment on the assumption of lognormality.

Lognormal distributions are expected when (1) many factors contribute to the variation among a set of samples (no one factor is a dominant determinant of the variation) and, (2) each factor tends to affect the sample value in an independent multiplicative way. The consensus of the Panel was that although lognormality is a reasonable starting assumption for the distribution of underlying single-sample residues, lognormality would not generally be expected for the distribution of residues found in composites. Even if the distribution of underlying single-sample residues were perfectly lognormal, the residue found in each composite is effectively a weighted arithmetic mean of some number of single samples. And a weighted arithmetic mean of lognormal samples is not expected to be lognormal itself.

The methods outlined by the Agency on statistical methods for composite data represent the classic approach to the analysis of data which is uncensored (e.g., no samples below limit of detection) and follows a single lognormal distribution. The estimates for the mean and variance of the lognormal distribution derive from a technique known as maximum likelihood estimation in which the probability of the data given the model is maximized. This technique is unbiased and correct in the situations where the assumptions are correct. The Agency is encouraged to continue along these lines with improvements outlined below.

Composite samples represent the weighted (by volume or surface area depending on the location of the pesticide) arithmetic mean of the individual components. If the individual components are lognormally distributed, the composite sample is not lognormally distributed. Hence, there is clearly the possibility of bias in the estimation of the mean and variance of the original distribution if the composite is assumed to also be lognormally distributed. This is illustrated in the first row of Table 1 (as prepared by FIFRA SAP member Dr. Christopher Portier) in which samples from a lognormal distribution are randomly generated on the computer, averaged (no weighting) in groups of 20 and then fit to a lognormal distribution. The numbers in columns 2 and 3 represent the expected mean and standard deviation of the original lognormal distribution based upon the method outlined by the Agency. It is clear that both the mean and standard deviation are overestimated and can be biased. The difference between the true single value distribution and the estimated distribution from assuming lognormality of the composite is given in Figure 1 (as prepared by FIFRA SAP member Dr. Christopher Portier). It is clear that, while the estimated distribution has greater mean, it's tail behavior is actually smaller than that of the original distribution possibly leading to some bias in estimations of the 95th percentiles. However, other methods are available, which could easily be used to alleviate this problem. One method is imputation in which computer generated distributions for the single sample residue levels are used to generate a distribution for the composite samples and this generated distribution, rather than using the theoretical lognormal distribution. While computer intensive, this method is likely to be more accurate than using the lognormal for the composite samples. To further illustrate this point, consider the individual sample data given in the background documents for carbaryl on apples, as prepared by FIFRA SAP member Dr. Dale Hattis (Table 2 below). In this illustration, the effect of progressive levels of truncation is to overestimate the geometric mean and underestimate the standard deviation. These two biases act in different directions to influence the expected number of samples likely to exceed particularly high residue levels, but this is not a generally desirable circumstance.

There is no apparent reason to exclude the censored data points from the analysis. In its most general form, maximum likelihood estimation can be defined as follows for uncensored, continuous data. If $f_{\mathsf{N}}(x)$ is the density function for the distribution from which the data (x) are derived and N represents the vector of parameters to be estimated in this model, then estimates for N are derived by solving the following formula:

$$\underset{\phi \in \Phi}{Max} \prod_{all\, x_i} f_{\phi}(x_i) \qquad (1)$$

where M describes the range of possible values for N. The method can easily be extended to include censored data by recognizing that the contribution of censored data to the likelihood is the probability that the observed concentration is below some known limit of detection, say L.

TABLE 1: Comparison of the Mean Estimates of m and s in a Lognormal Distribution with True Mean 1 and True Standard Deviation 1 Using Maximum Likelihood Methods Excluding Censored Data Points: (1) to Maximum Likelihood Estimation Including Censored Data Points (2) Based Upon 400 Simulated Data Sets. (C. Portier, F. Parham, F. Ye)

Limit of Detection (% samples lost)	Excluding Non-Detects and Using Equation (1)		Including Non-Detects and Using Equation (2)	
	Means	Variances	Means	Variances
0 (0%)	1.46	1.23	1.46	1.23
3 (8.7%)	1.51	1.08	1.46	1.25
4 (40.3%)	1.64	0.86	1.46	1.29
4.2 (47.2%)	1.67	0.80	1.45	1.28
4.5 (57.7%)	1.72	0.76	1.44	1.31
4.8 (66.6%)	1.78	0.72	1.44	1.33
5.0 (71.6%)	1.79	0.69	1.43	1.34
5.5 (81.7%)	1.87	0.64	1.42	1.38
6.0 (88.3%)	1.95	0.60	1.40	1.42
6.5 (92.8%)	2.01	0.53	1.40	1.41
6.6 (93.5%)	2.03	0.54	1.37	1.46

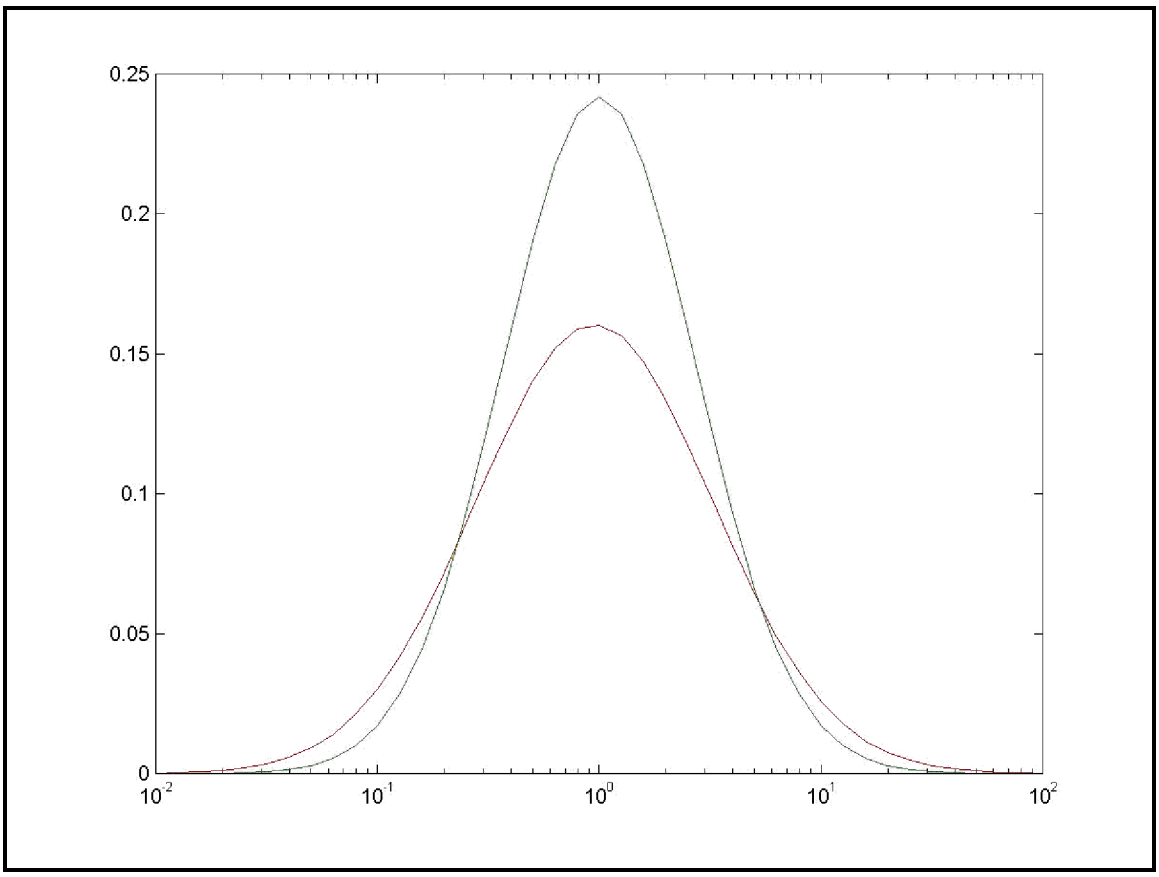


Figure 1: Comparison of a lognormal density with mean 1, standard deviation 1 with a lognormal with mean 1.46, standard deviation 1.23 (C. Portier)

Table 2. Effects of Truncation on Simple Method-of-Moments Calculations of Means and Standard Deviations (D. Hattis)

Data Set

Number of Points	Mean Log	Std Dev Log	Gmean	GSD	Est Arith Mean	Est 95th %tile	All
108	0.0923	0.2357	1.237	1.721	1.433	3.020	
Trunc .5	102	0.1275	0.1896	1.341	1.548	1.475	2.750
Trunc 1.0	74	0.2135	0.1430	1.635	1.390	1.726	2.810
Trunc 1.5	39	0.3208	0.1077	2.093	1.281	2.158	3.147
Trunc 2.0	20	0.4116	0.0652	2.580	1.162	2.609	3.303
Trunc 2.5	12	0.4465	0.0595	2.796	1.147	2.822	3.502

If $F_N(x)$ is the cumulative distribution function arising from the distribution defined by $f_N(x)$, estimates can be obtained by:

$$\underset{\phi \in \Phi}{Max} \prod_{all\, x_i} \left[f_\phi(x_i) I_i + F_\phi(L_i)(1 - I_i) \right] \quad (2)$$

where $I_i=1$ if the observed data value is not censored (a value was detected) and $I_i=0$ if the result is below the limit-of-detection given by L_i. This method is applicable to any underlying distribution function, not just the lognormal.

This simple modification is illustrated by the remaining entries in Table 1. A simulation study was conducted in which individual samples from the lognormal distribution with μ=1 and F=1 were randomly generated on the computer. Samples of size 20 were pooled (average concentration) to obtain a composite sample concentration. This was repeated until 100 samples were generated. Varying limits of detection were applied to these simulated data and the methods proposed by the Agency (remove non-detects and use formula (1) above) and the censored likelihood (using all data and formula (2) above) and estimates of μ and F were obtained. The entire process was repeated 400 times and the average values of μ and F were calculated. This gives an indication of the operating characteristics of the two methods for common samples and directly evaluates the degree of bias one might expect to see. It is clear from Table 1 that when the limit of detection censors a small portion of the distribution, both methods yield equivalent results. However, as the degree of censoring increases, the method which ignores non-detects becomes progressively worse with serious overestimation of the mean and underestimation of the standard deviation. In contrast, the method using the censored data remains effectively unbiased through the range of censoring levels. The difference is shown in Figure 2 (as prepared by FIFRA SAP member Dr. Christopher Portier) for the case where censoring occurs if the sample is less than 5. Again, even though the Agency's method has a higher mean, it's tail behavior is less than that of the uncensored method and could lead to underestimation of high exposure risks.

Note that the method based upon using the censored data (equation 2) continues to work effectively even for average sample sizes of only 12 detects and begins to fail when the average uncensored sample size approaches 7. In general, one would be ill-advised to use less than 5 uncensored values in any estimation. A better rule of thumb is to look at the ratio of the estimated parameters (N) and their standard deviations and avoid cases where this ratio is large.

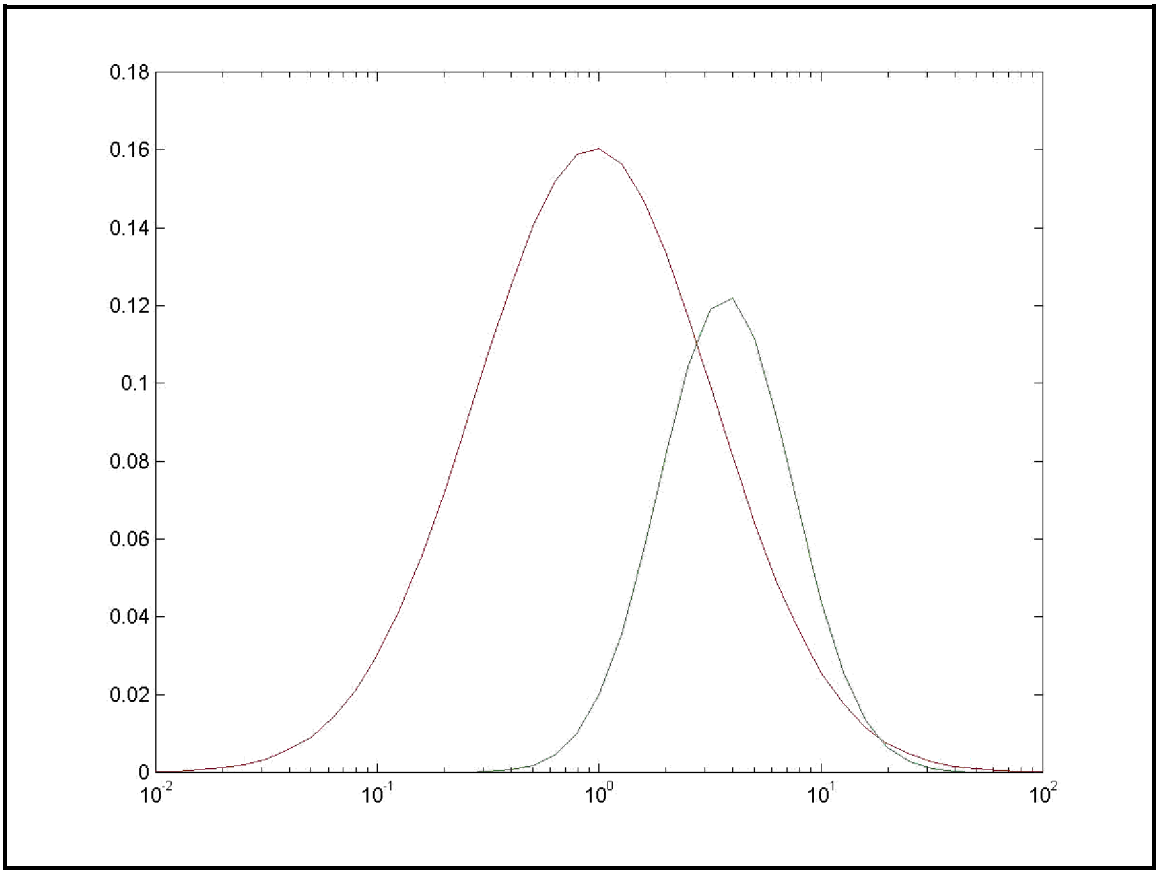


Figure 2: Comparison of a lognormal density with mean 1.46, standard deviation 1.23 with a lognormal with mean 1.79, standard deviation 0.69 (C. Portier)

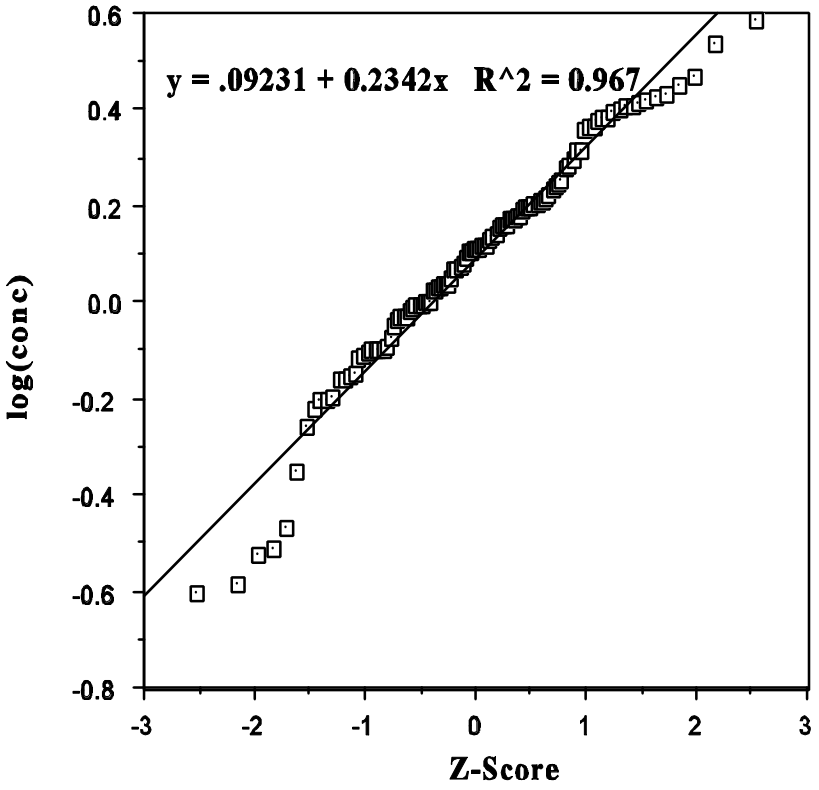
Several other issues should be considered in this type of evaluation. For example, the individual samples may arise from different distributions. In this case, you could get some degree of bimodality in the resulting data. Methods exist for stripping out multiple distributions from data, along the lines of equation 2, which could be used to perform a stepwise search of how many distributions may exist. Likelihood ratio tests or other applicable procedures could be used to decide if there is need for multiple densities in the evaluation. The mixture of distributions may also include a point mass at zero (some of the samples in the composite were never treated. Similar methods could be used for this case. To illustrate how a likelihood could be developed for such cases, consider the example below.

$$\underset{\phi_1\phi_2 \in \Phi}{Max} \prod_{allx_i} \left[f^1{}_{\phi_1}(x_i) I_i + F^1{}_{\phi 1}(L_i)(1-I_i) \right] \pi + \left[f^2{}_{\phi 2}(x_i) I_i + F^2{}_{\phi 2}(L_i)(1-I_i) \right] (1-\pi) \quad (3)$$

where f^1 and f^2 are two different densities (with matching cumulative density functions) with their own parameters to be estimated and π is an additional parameter describing what portion of the sample is ascribed to the density 1. Such procedures will require more data.

Finally, because of heterogeneity in growing conditions, pesticide use practices, and other factors, there is a possibility that some specific sources of variation will have a relatively large influence on residue levels. If one or two discrete circumstances have a large influence on residues, it is quite conceivable that distributions that are formed by mixing two or more lognormals would better describe the data. An example of composite residues that may be the result of multiple distributions of unit residues is given for 108 measurements of carbaryl on apples (Figure 3; D. Hattis). Here, the probability plot indicates significant divergence from a single, lognormal distribution.

Figure 3. Lognormal Probability Plot of 108 Sample Data Points for Carbaryl in Apples (Data of Harpenden, Apple 4995. [D. Hattis])



These difficulties seem to be well addressed in the analysis procedure offered by one of the public commenters during the meeting, Dr. Robert Sielken. The Panel did not have the opportunity to review the inner workings of the underlying software; however the general approach seems appropriate. The Panel recommends that Dr. Sielken publish the procedure and examples of its implementation in a peer-reviewed journal. Following this, the Agency actively explore the feasibility of using it or adapting it for the exposure estimation problems that were the focus of the session.

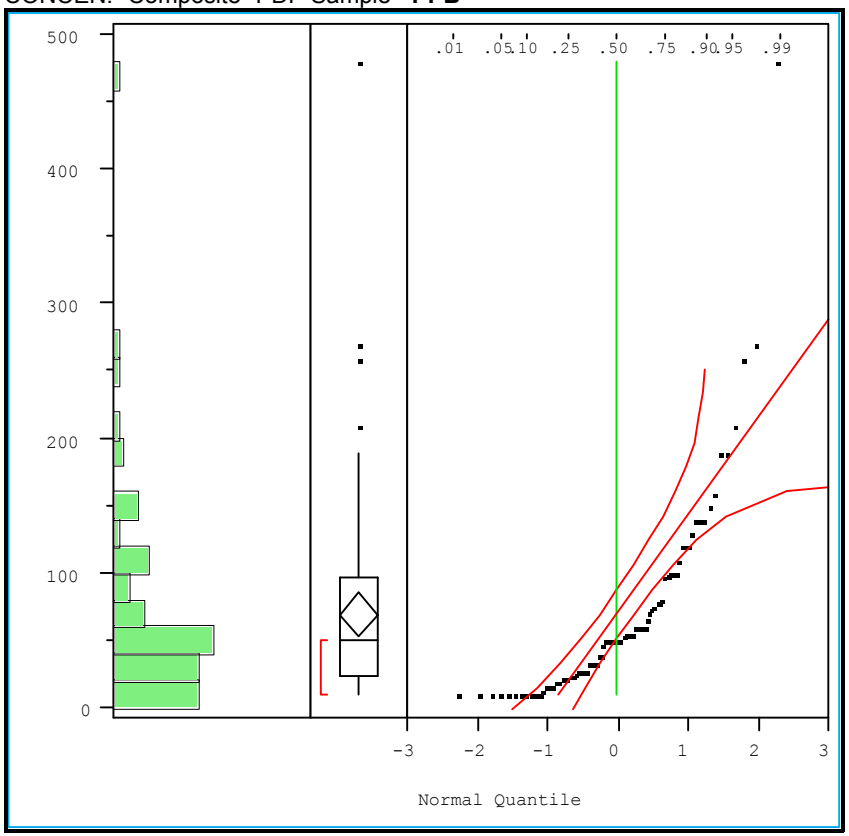
The Panel also provided several additional general comments in response to this question. The assumption that the lognormal statistical distribution actually fits all possible sources (i.e., PDP, FDA, registrant-field, processor, state market basket, land grant university bridging data sets, etc.) is a bit premature. An explanation of the total number of data sets submitted to a distribution fitting procedure is required. The inference or assumption deduced from these analyses is that the universe of all possible residue data sets are all represented by the lognormal statistical distribution. Without actually fitting a number of these data sets (e.g., n = 30), it is difficult to actually infer or adopt this assumption under sound statistical inference. This is presented in Figures 4-7 (as prepared by FIFRA SAP member Mark Whalon), an analysis of azinphos-methyl PDP data on apples and peaches. The lognormal transformation of the detect data improves the distribution of the data, but it fails a normal distribution test. The Agency should also include some background discussion of statistical sampling error as applied to the formation of composites and measurement of residue levels. This would help illuminate where the procedure might work and where it might be susceptible to errors or distortions of various kinds. In addition, when working with data sets and estimating parameter values, it may generally be more useful and appropriate to work with log-transformed data.

(2) The application of OPP's decomposition methodology calls for at least 30 "detects." This is done to assure that there is enough representation in the sample and that the extrapolation will cover the width of the distribution of single units. Although 30 detects is a practical rule for the application of the procedure, please comment on the consideration of other numbers as a practical rule of application.

The answer to this question depends on: (1) judgment concerning the desired degree of accuracy in the parameter estimates and (2) numerical experimentation in which different formulas for data acceptance are tested to evaluate estimates. As a general matter, the Panel believed that usable analyses should be possible based on data sets with many fewer than 30 "detects," and that no hard-and-fast numerical bright line should be specified. A factor that might be more significant, in the end, is the proportion of samples that have residues above the detection limit. Theoretically, it is possible to characterize the sampling distributions for uncertainty in statistical estimates associated with data sets of three or more samples if the data are a random representative sample and if the only significant source of uncertainty is random sampling error. In general, one should be cautious using even 30 samples since it may not eliminate uncertainty.

Figure 4. Pesticide Data Program 1993 - 1996: Azinphos-Methyl / **Peaches** Domestic Samples-raw data ; detects only (M. Whalon)

CONCEN: "Composite" PDP Sample - **PPB**

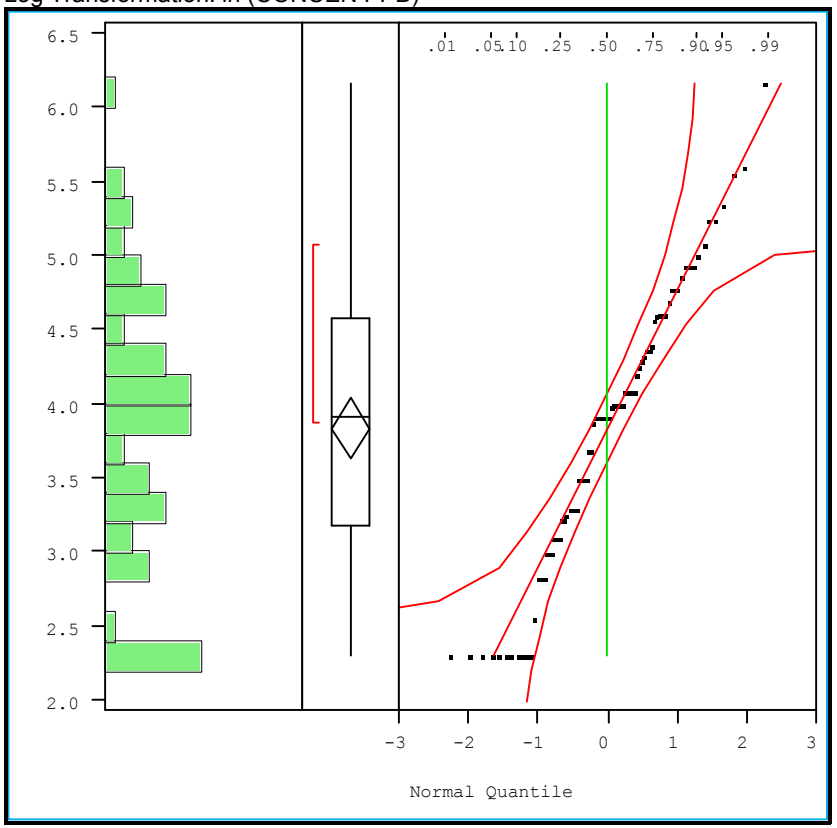


Mean	70.04878
Std Dev	72.84092
Std Error Mean	8.04393
Upper 95% Mean	86.05375
Lower 95% Mean	54.04381
N	82.00000

Test for Normality Shapiro-Wilk W Test

W	Prob<W
0.737930	0.0000

Log Transformation: ln (CONCEN PPB)



Mean	3.83396
Std Dev	0.93311
Std Error Mean	0.10304
Upper 95% Mean	4.03899
Lower 95% Mean	3.62894
N	82.00000

Test for Normality Shapiro-Wilk W Test

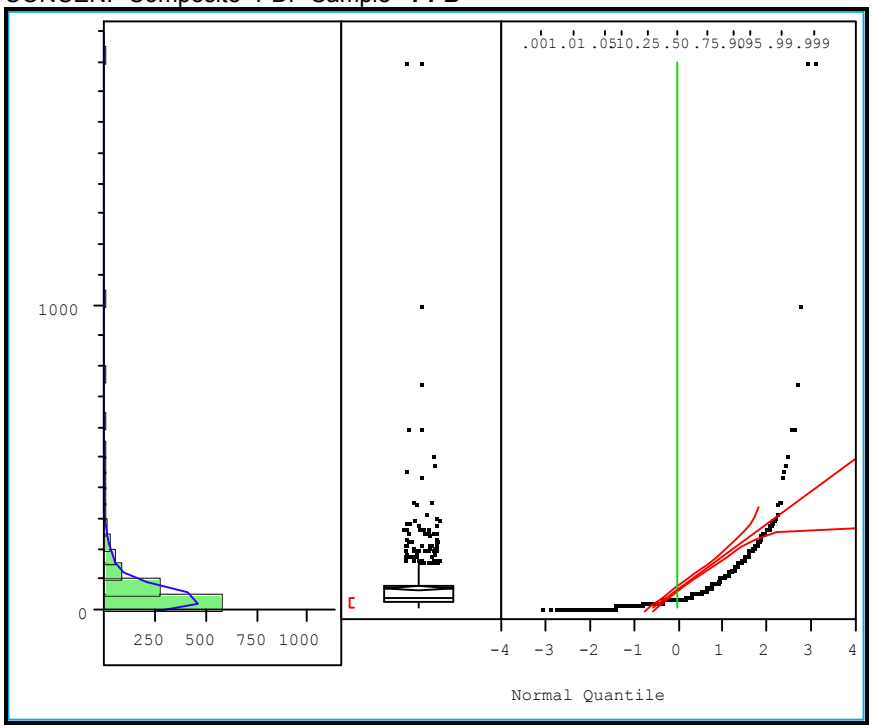
W	Prob<W
0.953278	0.0145

The " Test for Normality " tests that the distribution is normal If the p-value reported is less than .05 (or some other alpha), then you conclude that the distribution is not normal. If you conclude from these tests that the distribution is not normal, it is useful to use the Normal Quantile command in the check border menu to help assess the lack of normality in the distribution.

Upper 95% Mean and lower 95% Mean are 95% confidence limits about the mean. They define an interval which is very likely to contain the true population mean. If many random samples are drawn from the same population and each 95% confidence interval is determined, you expect 95% of the confidence intervals so computed to contain the true population mean. The upper and lower limits are computed as the sample mean, plus or minus a 97.25% Student's t value multiplied by the standard error of the mean.

Figure 5. Pesticide Data Program 1993 - 1996: Azinphos-Methyl / **Apples** Domestic Samples-raw data ; detects only (M. Whalon)

CONCEN: "Composite" PDP Sample - **PPB**

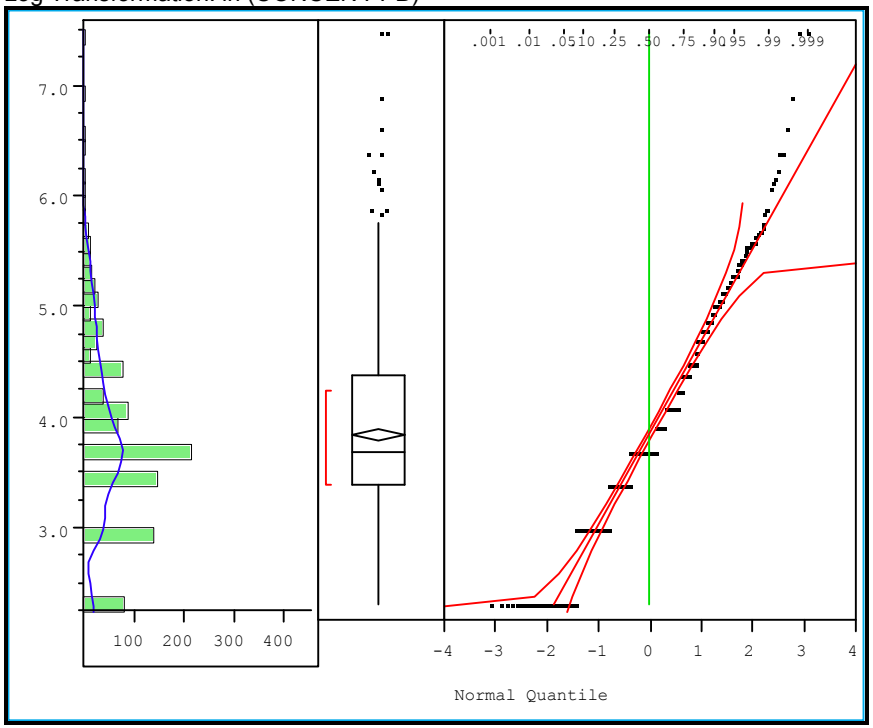


Mean	70.296
Std Dev	106.959
Std Error Mean	3.306
Upper 95% Mean	76.782
Lower 95% Mean	63.810
N	1047.000

Test for Normality: Shapiro-Wilk W Test

W	Prob<W
0.438479	0.0000

Log Transformation: ln (CONCEN PPB)



Mean	3.844
Std Dev	0.838
Std Error Mean	0.026
Upper 95% Mean	3.895
Lower 95% Mean	3.793
N	1047.000

Test for Normality: Shapiro-Wilk W Test

W	Prob<W
0.953989	0.0000

The "Test for Normality " tests that the distribution is normal If the p-value reported is less than .05 (or some other alpha), then you conclude that the distribution is not normal. If you conclude from these tests that the distribution is not normal, it is useful to use the Normal Quantile command in the check border menu to help assess the lack of normality in the distribution.

Upper 95% Mean and lower 95% Mean are 95% confidence limits about the mean. They define an interval which is very likely to contain the true population mean. If many random samples are drawn from the same population and each 95% confidence interval is determined, you expect 95% of the confidence intervals so computed to contain the true population mean. The upper and lower limits are computed as the sample mean, plus or minus a 97.25% Student's t value multiplied by the standard error of the mea

Figure 6. Example Lognormal Transformation of Azinphos-methyl PDP Detect data (M. Whalon)

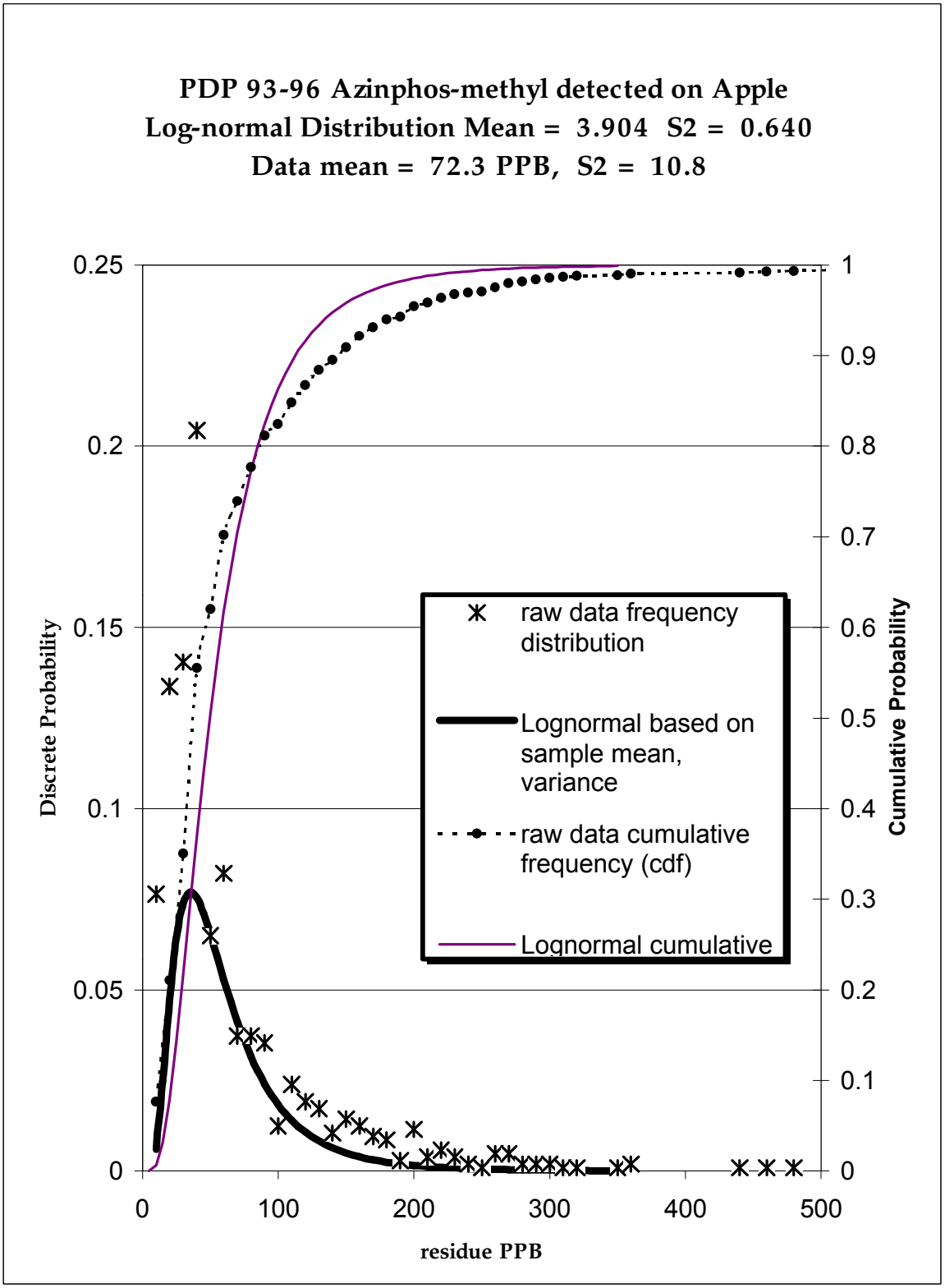
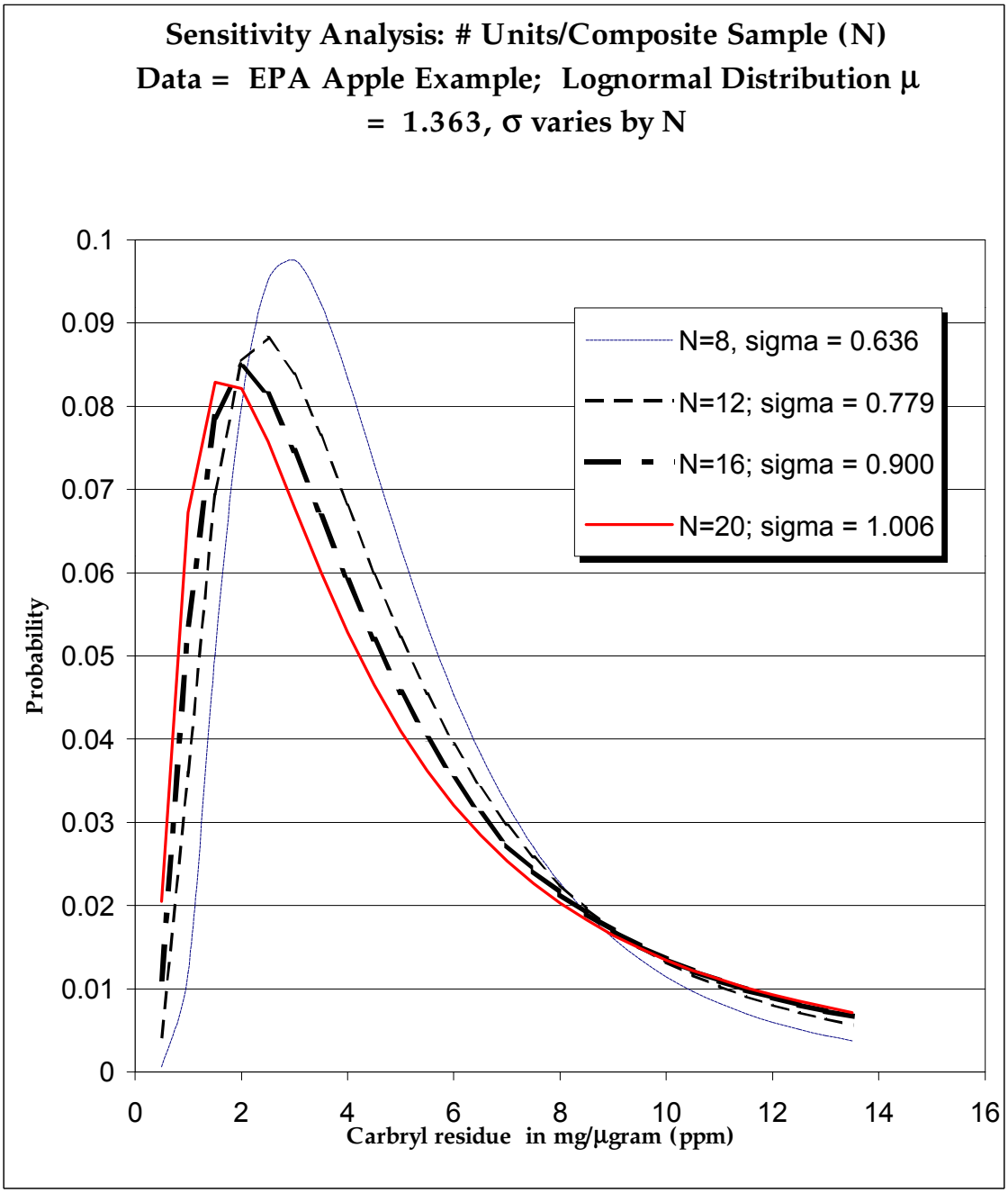


Figure 7. Sensitivity analysis of N: using the example data presented in the EPA background document (M. Whalon)



The general suggestion of the Panel was that OPP should seek thorough analysis to quantitatively characterize variability and uncertainty in its exposure estimates and draw whatever policy implications that seem to be indicated on a case by case basis, rather than attempt to set arbitrary bright lines for acceptable data quantity and type as inputs for analysis. The implications of small sample sizes can be evaluated by quantifying uncertainty in the estimates.

(3) The standard deviation within a composite cannot be greater than the standard deviation of the population of individual residues. Are there any circumstances where the standard deviation within a composite can be greater than the standard deviation of the population of individual residues?

The Panel had some difficulty with the wording of this question. Possible interpretations include: (1) is it possible for the standard deviation of inter-unit variability within a composite to be greater than the standard deviation of inter-unit variability for the entire population of units and (2) is it possible for the standard deviation of inter-composite variability to be greater than the standard deviation for the (unknown) population inter-unit variability.

In the first case, since standard deviations are estimated, it would be possible to obtain some situations in which the standard deviation for inter-unit variability among a finite number of randomly selected units within a composite might be greater than the true population standard deviation for inter-unit variability. This result would occur as an artifact of random sampling error. However, most likely, because of the method by which individual samples making up a composite are selected, it was judged unlikely that there would be a systematic bias in these estimates.

The answer to the second interpretation of the question was generally considered to be no, since the standard deviation of inter-composite variability is based upon averaging of inter-unit variability. In an extreme case, the standard deviations for these two might be similar only if the composites were comprised of homogeneous samples, which seems unrealistic for practical purposes.

In general, although some extreme cases were suggested based on non-random sampling in the formation of composites, it was difficult to imagine many real world cases where the standard deviation of residues on individual samples making up a composite would be expected to be greater than the standard deviation of a residues in a national sample of individual single-servings.

(4) OPP acknowledges that the collection of composite samples in the PDP protocol is not purely random; therefore, the decomposition procedure will produce an overestimation of the standard deviation of the lognormal distribution of residues on fruits and vegetables. Moreover, the overestimation of the standard deviation is accentuated to the degree that the collection of composition samples departs from pure randomness. The consequence of overestimating the standard deviation is that the high end of the estimates of residues in single units may exceed what occurs in reality. What criteria (if any) should be used to

establish an upper-bound on the amount of residue projected in a single unit to address the potential for overestimation of the standard deviation?

This question was somewhat vague. It was unclear as to whether the question was aimed at identifying an upper bound for residue concentrations among samples within a particular composite, or whether it pertained to estimating an upper bound over all samples in a population. For the most part, the Panel addressed the question as if it pertained to the latter interpretation.

The Panel felt that the use of all of the data and likelihoods based upon censored samples should alleviate much of the bias in the estimate of the standard deviations of the residues in single units. However, there was additional discussion of the role of outliers worth mentioning.

Outlier data points that are judged to result from errors in measurement or reporting should be excluded. However, beyond such errors, the Panel was generally skeptical of arbitrary rules that would exclude high-end data points, unless the rules could be based on some argument of physical impossibility. It was worth suggesting that a mass balance approach could be used to make a preliminary inference regarding the worst case concentration that might occur for a particular individual sample. In this mass balance approach, all of the mass of the residue observed in a composite would be assigned to a single sample and all other samples in the composite are assumed to have no residue. This would yield a worst case upper-bound concentration for samples within that composite, but not necessarily within the total population of individual samples. As an example, it would be reasonable to exclude any reported residue level that exceeds the level one could expect to result from spraying an undiluted sample of the pesticide as supplied by the manufacturer directly on the agricultural product, followed by immediate analysis.

There was some debate concerning the Agency's methods for dealing with high outlier residue data points. A 1998 Michigan study was designed to provide good laboratory practice measurement that links actual field use of pesticides on randomly sampled farms to farm gate, processing and handling residues. This study was conducted on eight commodities (i.e., apples, peaches, bilberries, tart cherries, grapes, asparagus, cucumber, and potato) and is available on the web (www.cips.msu.edu). When a high residue level was detected, a check routine was implemented. This routine checked all the analysis, calculations, sample chain of custody, and pesticide use and usage records. Several of the outlier data points were resolved as errors. If a data point was not resolvable by this routine, a second residue analysis was undertaken. Even with lognormal transformation, some of the PDP azinphos-methyl residue data points from apples and peaches were more than three standard deviations from the mean. Therefore, a PDP lab routine should be developed that identifies outlier residue values within the lab and back-checks the process leading to the outlier data point. This process may help resolve some of the outliers observed in the data.

If there were a sufficiently large number of single serving data sets, it is possible to evaluate the degree of overestimation of residues on single servings from composite data. Instead

of this process, PDP data could be examined and correspond recording of the number of composite outliers (i.e., stem and leaf displays with calculated confidence intervals) with abnormally high (i.e., greater than two or three standard deviations from the mean) composite residue means. PDP data do contain some residue values that are very high.

(5) OPP's methodology is sensitive to the number (N) of single units/servings of a commodity estimated to be in a composite sample. Please comment on how to estimate the number of single units/servings per composite sample. (Consider how to handle fruits for which a single unit is typically only a part of a unit of a commodity (e.g., a melon), or many different units (e.g., grapes) even though the single unit is smaller than the typical composite sample).

The numerical procedure used for analysis should mirror the sampling procedure used for construction of the composites. If the workers who construct the composite take three grapes each from 1000 different and independent crates of grates, then the right "N" for the composite calculation is 1000, even though people routinely eat more than three grapes per serving. Of course, when calculating the distribution of individual daily doses for different people, however, the distribution of quantities of grapes actually consumed should be used.

There is substantial literature on estimating sample sizes in survey sample methods. Some of this literature might contain ways to deal with circumstances such as that presented when the number of samples per composite is not a fixed single number but is actually a variable, depending on the variation in the size of individual fruits and vegetables of each type, and the variation in the sizes of the samples collected by the USDA surveyors.

Finally, it is possible to have dependencies in the sizes of agricultural products and consumption on the part of specific types of consumers. For example, children are said to tend to eat relatively small apples or a portion of larger apples. Where possible, and where known or suspected, such dependencies of exposures of particular types of consumers on particular characteristics of agricultural commodity samples should be preserved in OPP's exposure estimation methodology. Since most serving sizes are developed and standardized in the DEEM software (i.e., by age, population, ethnicity, etc.), the probabilistic estimates should already compensate for the population demographic variables.

(6) The decompositing procedure estimates the number of units in a PDP composite by dividing the weight of the composite by an average weight of an individual unit. The number of individual units in a composite will vary, depending upon the weight of each composite unit. Will differences in the number of individual units in a composite introduce substantial uncertainty?

The Panel concluded that the answer to this question can only be found, as before, by collecting data on how the number of units per composite varies among composites for specific commodities, followed by numerical experiments simulating effects on the calculations. Such an experiment was performed by one Panelist. The result was that uncertainty in the number of

units, such as from 13 to 16, in a composite, leads to approximately a 10 percent range of uncertainty in standard deviation of the inferred distribution for inter-unit variability using EPA's proposed method. The effect of this range of uncertainty on estimates of pesticide residues is case-specific. For example, the range of uncertainty associated with the 95th percentile of inter-unit variability in residues might be only one percent in one case but many factors in another case, depending on the nominal mean and standard deviation for inter-unit variability.

The question can also be answered relating surface area to volume ratio, area residues correlated to water weight or surface area of the fruit or vegetable. The response depends on the water solubility of the chemical in question, the movement of the chemical across the fruit or vegetable's leaf surface or cuticle and the other uptake routes (i.e., through roots, stems, reproductive tissues, etc.), together with the surface area of residue deposits. For systemic compounds, weight may be a better indicator of residue than surface area. For compounds that are not systemic, surface area probably better reflects residues.

NOTES

SAP Report No. 99-03C, May 27, 1999

REPORT:
FIFRA Scientific Advisory Panel Meeting, May 27, 1999, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Use of Watershed-derived Percent Crop Areas as a Refinement Tool in FQPA Drinking Water Exposure Assessments for Tolerance Reassessment

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

Christopher Portier, Ph.D
Chair
FIFRA/Scientific Advisory Panel
Date:_____

FEDERAL INSECTICIDE, FUNGICIDE, AND RODENTICIDE ACT
SCIENTIFIC ADVISORY PANEL MEETING
MAY 27, 1999

Session III -Use of Watershed-derived Percent Crop Areas as a Refinement Tool in FQPA Drinking Water Exposure Assessments for Tolerance Reassessment

PARTICIPANTS

Chair
Christopher Portier, Ph.D., National Institute of Environmental Health Sciences, Research Triangle Park, NC

FIFRA Scientific Advisory Panel Members
Ernest E. McConnell, DVM, Toxpath, Inc., Raleigh, NC
Herb Needleman, M.D., Professor of Psychiatry and Pediatrics, School of Medicine, University of Pittsburgh, Pittsburgh, PA

FQPA Science Review Board Members
Mr. Joel Blomquist, USGS/NAQWA, Baltimore, MD
Bernard Engel, Ph.D, Professor, Purdue University, West Lafayette, IN
William Gburek, Ph.D., Hydrologist, USDA/ARS, University Park, PA
Ms. Margaret Maizel, NCRI Chesapeake, Inc. Arlington, VA
Thomas Potter, Ph.D. Research Chemist, USDA/ARS, Tifton, GA
Ms. Gail P. Thelin, Geographer, USGS/NAWQA, Sacremento, CA
Harold Van Es, Ph.D., Professor, Cornell University, Ithaca, NY

Designated Federal Official
Mr. Paul Lewis, FIFRA Scientific Advisory Panel, Office of Prevention, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, DC

PUBLIC COMMENTERS

Oral statements were received from the following individuals:
Kim Winton, Ph.D., American Crop Protection Association
David Gustafson, Ph.D., Acetachlor Registration Partnership
Nick Poletika, Ph.D, Dow Agrosciences
Russell Jones, Ph.D., Rhone-Poulenc
Warner Phelps, Ph.D, Novartis Crop Protection

Written statements were received from:
Dr. Stuart Cohen, Environmental and Turf Services

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency regarding use of watershed-derived percent crop areas as a refinement tool in FQPA drinking water exposure assessments for tolerance reassessment. Advance public notice of the meeting was published in the *Federal Register* on May 5, 1999. The review was conducted in an open Panel meeting held in Arlington, Virginia, on May 27, 1999. The meeting was chaired by Christopher Portier, Ph.D., of the National Institute of Environmental Health Sciences, Research Triangle Park, NC. Mr. Paul Lewis served as the Designated Federal Official.

Agency scientists presented the "Proposed Methods for Determining Watershed-based Percent Crop Areas and Considerations for Applying Crop Area Adjustments to Surface Water Screening Models" to the FIFRA Scientific Advisory Panel. Sid Abel opened the session with an overview of the current approach for estimating pesticide concentrations in drinking water from surface water sources, outlined the Agency's recent work on the index reservoir, and identified critical assumptions of modeling and monitoring efforts. Dr. William R. Effland discussed the use of Geographical Information Systems (GIS) technology to estimate Percent Crop Areas (PCAs) from county-level agricultural census data and the application of the PCA to screening model estimates of pesticide concentrations in drinking water. Nelson Thurman compared the PCA-adjusted modeling results with monitoring data for corn-soybean herbicides in the midwest U.S. and described a preliminary comparison of screening model results with surface water monitoring data. Dr. Ian Kennedy employed 3-D graphical visualization tools to compare of monitoring data and PCA-adjusted modeling results for minor crops in the Central Valley, CA.

CHARGE

The specific issues to be addressed by the Panel are keyed to the background document entitled *Proposed Methods for Determining Watershed-Derived Percent Crop Areas and Considerations for Applying Crop Area Adjustments to Surface Water Screening Models, April 27, 1999* and are presented below.

1. Given preliminary comparisons between modeling and monitoring data, and the information presented regarding development of a Percent Crop Area (PCA), does the SAP think the PCA adjustment to PRZM/EXAMS modeling is a reasonable approach to obtain more accurate and appropriately conservative estimates of pesticide concentrations in surface water for screening evaluations of drinking water exposure? If the steps are not considered to be appropriately conservative for screening evaluations, does the SAP have any recommendations for how the EPA should use PCAs in its drinking water assessment process?

2. A GIS data processing method for calculating the PCA was presented. If we are able to resolve concerns regarding current model and monitoring inconsistencies, does the SAP think this

GIS procedure is an appropriate method to account for the portion of the watershed planted to the crop or crops of interest?

3. In estimating water concentrations for pesticides applied to multiple crops in a watershed, we modeled each crop separately, applied the maximum PCA to the modeled results, and then summed the outputs. For example, the model results for corn multiplied by the maximum PCA of 0.46 and the results for soybeans, multiplied by 0.41, were then summed to provide an estimate of pesticide concentrations for metolachlor use on corn and soybeans. Limitations to this approach may occur when the pesticide is used on multiple crops (such that the equivalent PCA is greater than 100%) or when the timing and/or rates of application vary for different uses. Can the SAP provide any recommendations for determining a reasonable assessment process that considers multiple uses of one or more pesticides within a watershed?

4. Evaluation of this project required PRZM/EXAMS modeling of selected crops with the preliminary version of the index reservoir scenario. Furthermore, the comparison relied on available reservoir and surface water monitoring data from limited sources (e.g., ARP, USGA NAWQA program). This evaluation produced variable results with some cases in which available monitoring data exceeded modeled results and other cases in which modeled data are conservative (i.e., clearly higher than available monitoring data). In the document overview, we have proposed steps to evaluate PRZM/EXAMS for inconsistencies as a screening model. What suggestions does the SAP have that could help us better understand these inconsistencies?

5. The PRZM developer has indicated the model has scale limitations when changing from field to watershed to basin scales. Possible sources of error when changing scales may be caused by using a single curve number for the entire basin, the hydraulic length, and simplifications for field scale processes that may not apply to the more complex basin scale. In addition, the application of the PCA in this report involves differing scales: PRZM is a field-scale model, the PCAs are generated from basin-scale areas (8-digit hydrologic units), and the index reservoir represents a watershed scale. Does the SAP have any recommendations for addressing the scale limitations of this model. Can the SAP suggest other watershed-based, mechanistic models that could estimate environmental concentrations of pesticides in drinking water?

6. The proposed application of percent crop areas for watershed modeling does not consider "percent crop treated" for one or several pesticides. Because pesticides in drinking water sources is a localized concern, a national percent crop treated estimate is not appropriate. At this time, data on percent crop treated at the farm are extremely limited, available only in New York and California. Does the SAP think it would be reasonable for the Agency to develop a method similar to the PCA approach to incorporate percent crop treated into the model refinements?

PANEL RECOMMENDATION

The Panel commends the OPP staff for its efforts in quantitatively evaluating pesticide exposure to drinking water. The Panel recognizes that this very difficult task is confined by the

limited resources available to the Agency for this purpose. The Panel also compliments the OPP staff on a very effective presentation of the research efforts.

The SAP is concerned about the general approach to screening pesticides. The proposed refinements, notably the percent crop area, are sensible improvements, but also raise new concerns, especially with minor-use crops. The use of the PRZM/EXAMS model will prevent the accurate evaluation of pesticide losses to reservoirs because of the inherent limitations of both the model and the input data. The model appeared to perform reasonably well with major crops in the Midwest and can be comfortably applied under those conditions. One SAP member suggested that, for cases where the proposed method cannot be accurately used, it is perhaps better to use a very simple screening approach. With chemicals of concern, subsequent evaluations may include sophisticated site-specific modeling efforts.

The Panel believes that more consideration should be given to low-cost, targeted monitoring, especially in the case of minor-use crops where modeling efforts tend to be imprecise. Instead of the current practices of widescale monitoring when pesticides are believed to pose great risk, the monitoring could be limited to a few watersheds that are deemed to be the most vulnerable. This approach is similar to the modeling efforts which also target the most vulnerable areas. In addition, some field experimentation may be executed using rainfall simulators to evaluate chemical losses under extreme rainfall conditions. This approach would provide adequate protection of the public's interest while preventing the unnecessary loss of a chemical to the agricultural industry.

Some SAP members expressed concern about pesticide degradation products which, although currently not included in the screening efforts, may still pose considerable human health risk. The Agency is encouraged to continue the evaluation of other models, as suggested by the July, 1998 SAP. Notably, the use of watershed-based models in a GIS environment should be given serious consideration. Perhaps the use of watershed-based regression models may be appropriate and should be considered, especially with increasing availability of high-quality monitoring data. However, regression models should be applied with caution, as extrapolation to areas/times/conditions beyond their development range may provide inaccurate estimates.

DETAILED RESPONSE TO THE CHARGE

1. Given preliminary comparisons between modeling and monitoring data, and the information presented regarding development of a Percent Crop Area (PCA), does the SAP think the PCA adjustment to PRZM/EXAMS modeling is a reasonable approach to obtain more accurate and appropriately conservative estimates of pesticide concentrations in surface water for screening evaluations of drinking water exposure? If the steps are not considered to be appropriately conservative for screening evaluations, does the SAP have any recommendations for how the EPA should use PCAs in its drinking water assessment process?

The concept of the PCA is both appropriate and reasonable for adjusting for the fact that chemicals are typically used on a fraction of the watershed land area. It provides a technically defensible approach to reduce estimates of acute and chronic pesticide exposures to levels similar to those found in monitoring data. Use of the maximum PCA appears to result in an appropriately conservative assessment for most chemicals for major-use compounds.

The proposed approach could also be improved in several areas. Models have generally been developed to predict the average observed values in most cases. For example, the Curve Number method is based on regression and thus it would be expected to estimate the average runoff value for a particular event. The value that might be observed could be much lower or much greater. This may explain the higher than anticipated pesticide concentrations for the two cases discussed in the documents.

Weather characteristics used within the model can have a significant impact on the results. More local weather should be used. One should also note that rainfall that is more conducive to runoff will occur only occasionally.

Even though the Panel agreed that the use of the PCA is a reasonable approach to estimate pesticide concentrations in surface water, the Panel did identify several limitations with the methodology. First, the PCA adjustment may result in an underprediction of chemical losses as some croplands contribute disproportionately to runoff. For example, if the watershed contains both row cropped land and forested land, the cropland would be expected to contribute more significantly to runoff and watershed discharge. The PCA adjustment would, in such case, not be conservative enough. A possible approach to addressing this issue is to perform a more in-depth analysis of the land use in the watershed, evaluate the relative extent and the types of other land uses, and make some appropriate adjustments.

Second, there was concern with the PCA derived from information on 8-digit HUC's (average size over 1,000 square miles) and county-based crop acreages. Most drinking water supplies are fed by smaller watersheds that may have different PCAs. It is advisable to evaluate the discrepancies between the proposed PCA and the PCA of actual drinking water reservoirs. PCA may result in biases with minor crops. The PCA factors one gets (especially for minor crops) for 8 digit watersheds versus 11 or 14 digit watersheds could be very different.

Third, the scale of the watershed does not allow certain factors such as landscape position, soil type, slope, etc., to be taken into account. This may be important with minor crops that are not grown on the typical soil type in the watershed. Some regions of the country have extended growing seasons with high-value crops being grown on limited acreage of the best soils with intensive management practices. Such fields are expected to contribute highly to pesticide losses.

Fourth, the proposed manner in which PCA's will be applied requires that a single maximum PCA be applied universally for a given pesticide across all regions and climatic zones. In contrast, PRZM-EXAMS will utilize region-specific soil and climatic data. Consistency will

make the process more technically defensible. There appear to be two possible approaches. In one, the Agency would choose a single "worst case" scenario for a pesticide-crop combination by the county with the maximum PCA for a given crop. In this case, the national PCA will determine the location of the "worst-case" PRZM-EXAMS simulations. Alternatively, simulations would be performed using climatic and region-specific soil and PCA data. For example, corn grown in the Piedmont would have as inputs to the model, the maximum PCA for the region and climatic and soil data. Midwest corn production would likewise be evaluated. The "worst-case" obtained from these simulations would be used in the screening process. The latter approach has the advantage of providing more realistic estimates of drinking water impacts and is therefore recommended to the Agency.

One Panel member raised concerns that some pesticides requiring further screening should have passed the PRZM/EXAMS screening, and this appeared to support the PCA approach. If there is a more rational reason, the PCA approach makes sense, but it is a mistake simply to use it as a way to make the screening methodology more conservative. A final comment was raised that the tolerance of false negatives should be defined in a statistical context.

2. A GIS data processing method for calculating the PCA was presented. If we are able to resolve concerns regarding current model and monitoring inconsistencies, does the SAP think this GIS procedure is an appropriate method to account for the portion of the watershed planted to the crop or crops of interest?

The GIS procedure as presented by the Agency is technically correct in principle. The Panel believes that the modeling and monitoring inconsistencies are due to limited data available for this procedure. Runoff potential of pesticides depends greatly on factors that are not part of the information database used for the calculation of PCA (e.g., soil type, position of fields in the landscape, etc.). This may be especially important in the case of minor crops where fields may be located in more vulnerable sites in a landscape (e.g., on steep slopes, or on bottomlands near streams). The 8-digit HUC and county-level statistics may not provide good estimates of actual PCAs.

The use of land use/land cover to adjust estimates of cropped areas within watersheds is a reasonable approach. Additional details of the approach need to be worked out. Some of the currently available national land use/land cover GIS data lump cropped areas and pasture together. This may be too coarse of a definition to be of much use in making adjustments to cropped areas within some watersheds. The minor crops will be most sensitive and adjustments will be most important for these crops. The Panel recommends that OPP staff specifically research possible biases of the PCA with minor-acreage crops.

3. In estimating water concentrations for pesticides applied to multiple crops in a watershed, we modeled each crop separately, applied the maximum PCA to the modeled results, and then summed the outputs. For example, the model results for corn multiplied by the maximum PCA of 0.46 and the results for soybeans, multiplied by 0.41, were then

summed to provide an estimate of pesticide concentrations for metolachlor use on corn and soybeans. Limitations to this approach may occur when the pesticide is used on multiple crops (such that the equivalent PCA is greater than 100%) or when the timing and/or rates of application vary for different uses. Can the SAP provide any recommendations for determining a reasonable assessment process that considers multiple uses of one or more pesticides within a watershed?

It appears inappropriate to sum the maximum PCA for one crop and that of another if these PCAs are not in fact based on the same watershed. A better approach is to determine the PCA based on the combined acreage of the multiple crops in any watershed. For example, the agricultural census county-level data for corn and soybeans would be added to determine the maximum PCA. The model estimates may then reasonably be adjusted using the aggregate PCA. If pesticide rates or timing of pesticide application vary by crop, then model estimates should be made for each crop separately, multiplied by the PCA for each crop, and then summed. The GIS could then be used to quickly derive combined percentages to identify the combined PCAs for the crop combinations of interest. OPP should perhaps "ground-truth" data by determining whether a particular set of crop combinations is "reasonable" for a given climatic and geographic region. Two used at the meeting were: 1) corn, soybeans, and wheat and; 2) corn, soybeans, and citrus. The former is a likely combination, while the latter is not. However, although corn, soybeans and wheat will likely co-occur in a watershed, it is highly unlikely that the relative areas of each crop will reflect the national maximum PCA. Crop co-occurrence should be evaluated for a given pesticide based on labeled use. This can be done by linking the PCA database and a labeled uses database. A "supervised" search will allow identification of a worst-case scenario, one where there is maximum potential chemical application rate for a realistic crop combination. In no case should crop coverage (via summation of PCAs) exceed 100 %. The development of such unrealistic scenarios will diminish the credibility of the "screening" effort.

Conversely, there is a concern about areas where double or triple cropping is used. This could be handled through the use of a PCA or by performing model simulations involving multiple crops in one growing season. The latter is preferred as it includes the effect of weather and soil conditions during different times of the year.

4. Evaluation of this project required PRZM/EXAMS modeling of selected crops with the preliminary version of the index reservoir scenario. Furthermore, the comparison relied on available reservoir and surface water monitoring data from limited sources (e.g., ARP, USGA NAWQA program). This evaluation produced variable results with some cases in which available monitoring data exceeded modeled results and other cases in which modeled data are conservative (i.e., clearly higher than available monitoring data). In the document overview, we have proposed steps to evaluate PRZM/EXAMS for inconsistencies as a screening model. What suggestions does the SAP have that could help us better understand these inconsistencies?

The inconsistencies could be the result of many sources, including inaccurate process representation in the model itself (preferential flow, groundwater discharge, etc.), scale issues (as identified in Question 5), quality of input data (soil, climate, chemical use, etc.), quality of monitoring data (sampling frequency, duration of study, etc.), and statistical data manipulation (e.g., arithmetic vs time-weighted averaging, UCL based on nonnormal data). Also, the PRZM/EXAMS modeling efforts are based on a small watershed, while some of the monitoring data may be collected from watersheds of a different size. Model estimates could be improved by using better data and more sophisticated tools, but these may require significantly more effort. A sensitivity analysis for model input parameters is advised. A first cut would be to look at soil half-life and adsorption data. Beyond this, hydraulic processes must be examined. In particular, the Agency needs to evaluate extreme events and define the appropriate weather "return" interval for use in the screening model. The Agency could carefully review published studies. There are numerous investigations described in the peer-reviewed literature. Whenever possible, the mass of pesticide discharged in runoff in the model and in published (rainfall simulation) studies should be compared to assist in evaluating model limitations.

The rainfall data for the cases in which observed values exceeded the modeled results may be responsible. The soil information for these locations should also be examined in these cases. The Salem, Illinois, soils are susceptible to runoff and a restricting layer is present in many of these soils. Combined with unusual rainfall, this could result in the observed pesticide concentrations in runoff.

The soil and water assessment technology model (SWAT) could be considered as a possible alternative. A GIS link with ArcView is available to facilitate the use of SWAT. The agricultural policy/environmental extender model (APEX) may be useful for smaller watersheds. Again, one should ask whether the purpose is screening or modeling.

Perhaps a more fundamental issue is the basic incompatibility between the modeling effort and the "real" world. To adequately model the real situation, considerably more effort needs to be expended. At this stage, one moves from a screening approach to attempting to model the actual environmental conditions. Ultimately, this should perhaps be pursued. If a screening approach is used that employs simplified process representations, then the results should not be expected to approximate actual monitoring data and should in nearly all cases exceed them. OPP should carefully evaluate all cases where model results do not exceed monitoring results.

The modeling results compared fairly well against monitoring data for major-use crops, especially with the Acetochlor Registration Partnership, which included many watersheds of similar size to the model watershed. This suggests that the PRZM/EXAMS approach is satisfactory for crops with high PCAs and for conditions of the midwestern USA. The results of minor-use crops in other geographic regions of the country were less than satisfactory, and OPP is advised to further investigate possible sources of error for those cases.

One should keep in mind that models have typically been developed to predict the mean value of parameters of interest. Thus, it is to be expected that a model would underpredict observed values in some instances and overpredict values in other instances. For the current approaches, the parameterization of the model to obtain conservative results normally results in overprediction of observed data in almost all instances. After exploring the cases in more detail in which the observed pesticide concentrations were greater than those predicted, an appropriate decision about the model parameterization approach can be made. If the unanticipated observed results cannot be explained, then a modification of the parameterization approach might be considered. One needs to be careful of the distinction between screening and modeling. A modeling approach will use much more data, be time consuming and thus more expensive. Modeling might be used as a Tier 3 tool.

5. The PRZM developer has indicated the model has scale limitations when changing from field to watershed to basin scales. Possible sources of error when changing scales may be caused by using a single curve number for the entire basin, the hydraulic length, and simplifications for field scale processes that may not apply to the more complex basin scale. In addition, the application of the PCA in this report involves differing scales: PRZM is a field-scale model, the PCAs are generated from basin-scale areas (8-digit hydrologic units), and the index reservoir represents a watershed scale. Does the SAP have any recommendations for addressing the scale limitations of this model. Can the SAP suggest other watershed-based, mechanistic models that could estimate environmental concentrations of pesticides in drinking water?

The scale issues are very profound and of concern to the Panel. Similarly, it should be noted that process-representation needs to be optimal. For example, in tile-drained fields, preferential flow is an important process affecting pesticide loading to streams. Also, PRZM does not account for groundwater discharge to streams, which may be important with chemicals that readily leach and are used on permeable soils near streams. The 8-digit HUC is generally not representative of drinking water reservoirs.

As noted above in response to Question 4, both the SAP and OPP recognize the serious limitations associated with the scale incompatibilities. These can be addressed but would require considerable investment of resources. OPP must address the purpose of watershed or basin-scale modeling. PRZM is not the answer if the Agency is seeking to develop one representative watershed or basin configuration. Also, the assumption of a uniform basin is not workable. Recognizing this, there are two options: attempt to incorporate non-uniformity through the use of a different model (e.g., ANNAGNPS), or use a simplistic approach with PRZM and account for uncertainty through the use of safety factors.

6. The proposed application of percent crop areas for watershed modeling does not consider "percent crop treated" for one or several pesticides. Because pesticides in drinking water sources is a localized concern, a national percent crop treated estimate is not appropriate. At this time, data on percent crop treated at the farm are extremely

limited, available only in New York and California. Does the SAP think it would be reasonable for the Agency to develop a method similar to the PCA approach to incorporate percent crop treated into the model refinements?

The Panel agreed that this is an important issue and needs to be pursued by the Agency. It appears advisable to analyze the New York and California data to evaluate the extent of discrepancy between "percent crop area" and "percent crop treated" and then make an educated decision on how to handle this issue. Similarly, OPP may employ use data collected by the agricultural chemical industry and further make use of information from experienced extension agents in making reasonable estimates. USDA-NRI data and the USDA National Agricultural Statistics Service internet site are readily available sources. Evaluations should include sensitivity analyses. Relatively high uncertainties may be encountered for chemicals which are applied to less than 10 % of the cropped area. Additional insights might be gained by comparing the compounds and their use in different geographic areas (e.g., grapes in California and New York).

While the majority of the Panel agreed that percent crop treated should be considered in model refinements, one Panel member disagreed, commenting that the use of percent crop treated data moves beyond the original intent of the PRZM screening approach. Use of this data would increase specificity of the technique. Thus, the Agency should not consider using this data.

Some pesticides are used on a relatively small fraction of the acreage of a certain crop, or may be minimized based on Integrated Pest Management procedures. Assuming that all crop land is actually treated, however, may appropriately represent a worst-case scenario for screening purposes. Conversely, it is also possible that the same pesticide is used repeatedly on the same crop during a growing season, which should also be accounted for. This is best done by simulating multiple applications in the model runs, rather than by making a correction afterwards.