SAP Report No. 2001-04
March 21, 2001

FIFRA Scientific Advisory Panel Meeting
September 27, 2000
Held at the Sheraton Crystal City Hotel, Arlington, Virginia

REPORT

SESSION II END POINT SELECTION AND DETERMINATION OF RELATIVE POTENCY IN CUMULATIVE HAZARD ASSESSMENT: A PILOT STUDY OF ORGANOPHOSPHORUS PESTICIDE CHEMICALS

Ms. Olga Odiott
Designated Federal Official
FIFRA/Scientific Advisory Panel
Date:_____

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

NOTICE

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

The FIFRA SAP was established under the provisions of FIFRA, as amended by the Food Quality Protection Act (FQPA) of 1996, to provide advice, information, and recommendations to the 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 FIFRA SAP on an ad-hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Larry Dorsey, SAP Executive Secretary, via e-mail at dorsey.larry@.epa.gov.

TABLE OF CONTENTS

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
September 27, 2000

PARTICIPANTS

FIFRA Scientific Advisory Panel Session Chair
Ronald J. Kendall, Ph.D., Director and Professor, Texas Tech University/ Texas Tech University Health Sciences Center, Reese Center, 1207 Gilbert Drive, Building 555, Lubbock, Texas 79416

FIFRA Scientific Advisory Panel
Fumio Matsumura, Ph.D., Director, Institute of Toxicology and Environmental Health, ITEH Building, University of California, Davis, CA 95616

Stephen M. Roberts, Ph.D., Professor and Program Director, University of Florida
Center for Environmental & Human Toxicology, Building 471 Mowry Road
Gainesville, Florida 32611

FQPA Science Review Board Members
Rory Conolly, Sc.D., Senior Scientist, Chemical Industry Institute of Toxicology, 6 Davis Drive Research Triangle Park, North Carolina 27709

Pat Durkin, Ph.D., 5100 Highbridge Street, Building 42C, Fayetteville, NY 13066

Dale Hattis, Ph.D., George Perkins Marsh Institute**,** Clark University**,** Worchester, MA

Ernest McConnell, Ph.D., President, Toxpath Inc., 3028 Ethan Lane, Raleigh, NC 27613

Peter Macdonald, D. Phil., Professor of Mathematics and Statistics, Department of Mathematics and Statistics, McMaster University, 1208 Main Street West, Hamilton, Ontario, Canada L8S4K1

Carey Pope, Ph.D., Department of Physiological Sciences, 264 McElroy Hall, Oklahoma State University, Stillwater, Oklahoma 74078

Nu-May Ruby Reed, Ph.D., Staff Toxicologist, California Environmental Protection Agency, Department of Pesticide Regulation, 830 K Street, Sacramento, California 95814-3510

Lester Sultatos, Ph.D., Department of Pharmacology and Physiology, New Jersey Medical School, 185 South Orange Avenue, Newark, New Jersey 07103

Designated Federal Official
Ms. Olga Odiott FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, DC

PUBLIC COMMENTERS

Oral statements were made by:
Christopher Wilkinson, Ph.D., from Jellinek, Schwartz, and Connolly Inc.; representing the American Crop Protection Association (ACPA)

Written statements were received from:
Angelina J. Duggan, Ph.D. , 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 the end point selection and determination of relative potency in cumulative hazard assessment of organophosphorus pesticide chemicals. Advance notice of the meeting was published in the *Federal Register* on September 5, 2000. The review was conducted in an open Panel meeting held in Arlington, Virginia, on September 26-29, 2000. The meeting was chaired by Ronald J. Kendall, Ph.D. Ms. Olga Odiott served as the Designated Federal Official.

The Agency presented an approach to assess and select common mechanism endpoints for accumulating hazard and a method for determining relative potencies for organophosphate pesticides. The relative potencies were used in a cumulative risk assessment of multiple organophosphate pesticides. The Science Advisory Panel was asked to advise on advantages and disadvantages of the method used to assess cumulative hazard and rank the pesticides.

CHARGE

The specific issues addressed by the Panel are keyed to the background document, "End Point Selection and Determination of Relative Potency in Cumulative Hazard Assessment: A Pilot Study of Organophosphorus Pesticide Chemicals", dated September 1, 2000, and are presented as follows.

SESSION II END POINT SELECTION AND DETERMINATION OF RELATIVE POTENCY IN CUMULATIVE HAZARD ASSESSMENT: A PILOT STUDY OF ORGANOPHOSPHORUS PESTICIDE CHEMICALS

Question 1 For most compartment and sex groupings, there were one or more chemicals for which multiple studies could be used to calculate an ED 50. (See Figures 2, 3, 5, 6, 8, and 9.) Please comment on the criteria OPP used to select the "representative study" from among the available studies available for a specific chemical to calculate the ED50 for that chemical, compartment, and sex.

Question 2 There will be situations for some chemicals in which data are lacking for the critical measurements, in a certain species or sex, or for a certain route of exposure. The lack of data may be because critical measurement(s) simply were not measured or because data are considered to be of poor quality.

Q2.1 We would like the panel's view on the use of surrogate data as a substitute

for the lack of appropriate data. To what extent should surrogate information be used to determine a chemical's relative potency?

Q2.2 How should situations be handled where an ED_{50} can be determined for many of the chemical members but cannot be determined for a few members? We would like the Panel's view on the use of NOAELs as substitutions for ED_{50}s for points of comparison.

Question 3 We would like the Panel's view on the relative importance of the factors discussed in the paper for selecting an index compound.

Question 4 We would like the panel to comment on the log dose-probit analysis used to extrapolate the ED_{50}s for the chemicals evaluated in this pilot.

Question 5 For this group of chemicals, OPP has sufficient data to calculate relative potency factors by the oral route for six different compartment/sex groups. Relative potency factors could be calculated by use of a single compartment/sex or by compiling data across compartment/sex groups.

Q5.1 If the Panel favors a single compartment/sex, please comment on the criteria that should guide the choice of a compartment/sex group.

Q5.2 It is proposed in this pilot analysis, that data could be compiled across different studies to provide more confidence in the determination of relative potency. In establishing an effective dose (e.g., ED_{50}), to what extent should one compile data for each chemical of interest within or across different measures and/or studies? What are important criteria to consider when compiling data?

Question 6 Dose addition is considered an appropriate default approach to cumulative risk assessment. The mathematical definition of dose-addition requires a constant proportionality between the effectiveness of the chemicals being considered. It is anticipated that extensive dose-response data will not be available for many chemicals. Please comment on the approach taken to evaluate parallel dose-response curves. Please comment on how rigorous an analysis is needed to evaluate the assumption of parallel dose-response curves.

Question 7 How does one handle a response for a chemical that displays a different slope (i.e. an outlier)? Examples were demonstrated in this pilot analysis where one or a few chemicals of the common mechanism group exhibited pronounced species, sex, or compartment differences from the majority of chemicals.

SUMMARY OF PANEL RECOMMENDATIONS

The Panel members concluded that while the criteria used for selecting a representative study appear reasonable, the Agency's background document is not clear on how these criteria were used in the selection process.

The Panel noted that when the compounds are ranked on the basis of their mean ED_{50}s, rankings do not change much when compared to the rankings based on the representative ED_{50}s. Therefore it is not readily apparent why a representative study is necessary, rather than simply using the mean ED_{50}s to rank the potencies.

There would be much greater confidence in the ED_{50} if it were derived from several, relatively consistent studies as opposed to a single study, without benefit of confirmation by other studies.

The use of surrogate data is reasonable when the data utilized for the extrapolation are of high quality, and there is an adequate understanding of the relationships among the data that are utilized for the estimation of the surrogate data. It is imperative that the Agency be transparent in why and how the surrogate data are chosen. There should be a discussion as to the “degree of confidence” or “level of uncertainty” the Agency has in the surrogate data.

The Panel was of the opinion that since the no-observed-adverse-effect-levels (NOAELs) will be usually substantially lower than their corresponding ED_{50}s, it could be misleading to mix them in the same analysis. The decision to use the NOAEL as a surrogate for the ED^{50} should be made on a case-by-case basis, with the overarching goal of selecting a value that minimizes the overall uncertainty in the cumulative risk assessment.

The most important considerations in selecting an index compound are that it acts toxicologically as purely as possible by the common mechanism defining the group, and that quantitative data for assessing potency be available for as many routes of exposure, genders, species and strains as possible that might contribute relative potency data to the analysis of the analysis of the relative potencies of other chemicals.

The Panel commented on the importance of modeling the relative potency factor (RPF) after a chemical that has the best and most complete data for the common endpoint(s). Having the best data for the index chemical could ensure that the uncertainties in the next step of the risk assessment are minimized.

The Panel noted that the statistical technique used in the pilot study is a probit transformation, not a "probit analysis" as described in the Agency's background document. To differentiate, the methodology used for the pilot study will be referred to as "probit transformation".

The Panel indicated that for any statistical analysis that involves fitting a model to data (preferably data on individuals), statistical tests indicating that the model provides an acceptable fit to these data should be given. Greater use of uncertainty analysis should be made. Quantitative uncertainty analyses should be developed to the extent possible. This would help with the overall evaluation of the statistical modeling approach being used.

The Agency should use individual animal data, reconsider the selection of the probit model and discuss why a particular model is selected and how well the model fits the data both conceptually and statistically.

The Panel generally felt that combining datasets across compartments and sex groups should be avoided. While female rats tend to be more sensitive to some of the toxicants in this class, it is unclear that this would always be the case. Thus, selection of males or females as a rule may not be appropriate.

In combining different studies, only studies of acceptable quality should be used. Studies that are obvious "outliers" should not be used. However, there needs to be a transparent explanation as to why a given study was deleted from the compilation of studies.

The Panel commended the Agency for developing such a sizable database. Besides being useful for eventually establishing RPFs for addressing cumulative risk of organophosphates, this database could be useful for furthering the understanding of various aspects of the toxicity of organophosphate pesticides. With this database, further investigation into the need for an uncertainty factor for subchronic-to-chronic extrapolation may now be possible.

The Panel stated that a rank order correlation is not a meaningful statistical test for parallelism, and that the underlying experiments were not designed to address the question of parallelism. More complex, biologically motivated approaches, such as PBPK modeling, would not require some of the constraining assumptions inherent in the probit approach, such as parallelism of dose-response curves.

The Panel noted that approaches that are more mechanism based, such a PBPK/PD modeling, would remove the need for constraining assumptions such as parallelism of dose-response curves.

When using the probit approach, uncertainty analyses, preferably quantitative uncertainty analyses, should be used to evaluate the consequences for the analysis of alternative approaches to the handling of outliers.

The assumption of parallelism is based on the further assumption that the animals used in the bioassays are randomly taken from the same population. Pesticide bioassays on file with the Agency have been conducted at different times, in different locations and may differ in other ways that are not readily apparent. Thus, the failure to demonstrate parallelism may have a more

complex interpretation than simply suggesting that the mechanisms of action are different.

Substantial reliance should be placed on what we know about the mechanism of action. The determination of whether or not to include the compound in the assessment of cumulative risk should be guided largely by the confidence in the knowledge that the mechanisms of action for the chemicals in the presumed class are properly understood.

Perhaps most importantly, the issue of outliers based on a failure to estimate parallel slopes may be incidental and irrelevant if the analysis uses an inappropriate dose-response model.

DETAILED RESPONSES TO THE CHARGE

Question 1 For most compartment and sex groupings, there were one or more chemicals for which multiple studies could be used to calculate an ED_{50}. (See Figures 2, 3, 5, 6, 8, and 9.) Please comment on the criteria OPP used to select the "representative study" from among the available studies available for a specific chemical to calculate the ED_{50} for that chemical, compartment, and sex.

The criteria used for selecting a representative study were the achievement of steady-state conditions, and the goodness of fit of the dose-response curve. However, the background document is not clear on how these criteria were applied. Indicators of "goodness-of-fit" for the dose response curves are not presented. How the fits were evaluated is an important issue since the document states that dose response curves contained only two to four data points (p. 62). The document is not clear on how consistency was evaluated. (the expression "dose-response curve consistent with the other studies" is a somewhat vague statement).

A variety of opinions were expressed by Panel members regarding the use of representative studies:

> In general, both of these criteria appear reasonable for selecting a representative study. Furthermore, examination of the ED_{50} values from the representative studies relative to the means for all studies generally indicates that the values from representative studies fall within the range of variability of the group as a whole. Thus, the studies selected could reasonably be considered "representative". In some cases, however, (e.g., Figure 6, chemical M), the representative study ED_{50} falls outside of the range of variability around the mean. In addition, while the criteria seem reasonable, it is unclear from the text of the draft document exactly how these criteria were used in the selection process.
>
> The representative studies that were chosen seem to adequately represent the other studies when they fall within one standard deviation of the mean value. However, when the ED_{50} from the representative study is more than a standard deviation away from the mean ED_{50}

an argument can be made that it is not representative of the group. And finally, it should be noted that when the compounds are ranked on the basis of their mean ED_{50} s, rankings do not change that much, with one or two exceptions, compared to the rankings based on the representative ED_{50} s. It is not readily apparent, therefore, why a representative study is necessary, rather than simply using the mean ED_{50} s to rank the potencies.

The criteria are reasonable, but it is not clear that the analysis is best served by the representative data approach. There would be much greater confidence in the ED_{50} if it were derived from several, relatively consistent studies as opposed to a single study, without benefit of confirmation by other studies. With the representative study approach, ED_{50} s derived from either situation would be considered equally valid, which is not necessarily the case. A Panel member's proposal to make better use of the data in determining the steady state inhibition is attractive. However, EPA's presentation of the examples of data sets available for analysis suggests that this might not work well in all situations.

The multiple study approach should be used when the data are amenable. For example, we know a lot about the pharmacokinetics of organophosphates (OPs) and classical pharmacokinetic studies in rats are typically required as part of the registration process. Rather than plotting ED_{50} values for various time points, a plausible response/dose/duration model could be used. Whether or not the temporal relationship apparent in the toxicity studies is consistent with the known kinetic parameters of the compounds should be assessed. It should be noted however, that with any large data set there would be "outliers" that need to be censored from such a grouping.

A Panel member suggested the following approach: make a preliminary screen of all the available data for each cholinesterase inhibitor, and then, analyze all the studies in different animal strains, genders, and at various time points, that can reasonably contribute valid information. Duration of exposure/dosing should be modeled as an explicit variable, and available results projected to an asymptotic "steady state" at essentially infinite time (or the maximum lifespan of the animal if there is expected to be appreciably incomplete attainment of steady state by then for some cases). The following function (subject to checking with actual data) would have the difference between the potency observed at time t and the steady state potency decline exponentially:

$$\text{Potency}_{\text{observed at time t}} = \text{Potency}_{\text{steady state}} \left(1 - e^{-kt}\right)$$

where the rate constant k is to be estimated from the data, and potency can be expressed in terms of the reciprocal of the dose that gives a particular percentage inhibition of the cholinesterase in question, or some other measure (such as "inhibition hits/dose rate"). This reflects a mechanistically appropriate expectation that "steady state" is approached, but never exactly reached, and allows adjustment and use of data from time points where inhibition is clearly less than that attained with further dosing. Using more than one study and time point data has two benefits—it will allow a more robust estimation of the central value for steady state potency for

the chemical (within a compartment and gender), and it will allow a better estimation of the uncertainty associated with the available data, crossing multiple time points, strains of animal, experimental measurement systems for cholinesterase levels, etc.) Determination of uncertainties in the relative potency estimates is important for later estimation of the uncertainties in estimated risks from combined exposures to multiple agents.

The Panel noted some errors in the tables: Inspection of Tables 1-3 indicates that Tables 1a, 1b, and 2b contain discrepancies. For example, in Table 1a many ED50s for the representative studies seem to have been incorrectly calculated from the logED50s. Some discrepancies are large, such as chemical O, where the antilog of 1.76 is 57.54, not 158.79. Other discrepancies are smaller, such as chemical I, where the antilog of –0.37 is 0.43, not 0.36. In addition, certain values seem to be reversed for chemicals H and P. Similar errors, although not as many, were found in Tables 1b and 2b. No mistakes were found in Tables 2a, 3a, and 3b. It is not clear the extent to which these discrepancies affect the rankings of potency. Given that these mistakes do not essentially alter the document in a meaningful manner, the criteria utilized seem to be appropriate.

Question 2 There will be situations for some chemicals in which data are lacking for the critical measurements, in a certain species or sex, or for a certain route of exposure. The lack of data may be because critical measurement(s) simply were not measured or because data are considered to be of poor quality.

Q2.1 We would like the panel's view on the use of surrogate data as a substitute for the lack of appropriate data. To what extent should surrogate information be used to determine a chemical's relative potency?

The concept of using surrogate data with its inherent uncertainties is not new to risk assessment. The key has always been in balancing the weight of support for using surrogates, the uncertainties that are introduced, and the consequence of not being able to estimate the risk. The support and extent of using surrogate data would require a careful evaluation of the evidence for a sex-, species-, or route-specific difference in sensitivity. The specific case in point is the use of surrogate data of cholinesterase inhibition (ChEI) endpoints for when "*data are lacking for the critical measurements, in certain species or sex, or for a certain route of exposure*." Some of the considerations are:

Surrogate between the males and females: The overall pattern from the pilot study (Figures 4, 7, 10) shows that females rats are equal or slightly more sensitive than male rats through the oral route of exposure. This general pattern could be used as a default adjustment for a sex-specific sensitivity. Specifically, for the missing data in females for chemical D (plama ChEI), and Chemical P (red blood cell (RBC) ChEI), and Chemical D (brain ChEI), male relative potency factors could also be considered as surrogates for the females, unless there is evidence to indicate otherwise.

Surrogate between routes: A route-to-route extrapolation of relative potency factors (RPFs) may be necessary when dose-response data are lacking (e.g., as shown in the pilot study for dermal or inhalation routes) for a major route of human exposures. A prudent approach for a route-to-route extrapolation of a dose-response relationship for systemic effects would include considerations of all available pharmacokinetic information (e.g., route-specific absorptions, the metabolic capacities at the site of entry).

Surrogate between species: Further studies based on the existing pilot study data would be needed for determining appropriate surrogates for the missing chemical-specific data for Chemical Y (RBC ChEI) and Chemical F and J (brain ChEI). If the database is sufficiently rich, it may be possible to use RPFs from another species to fill these data gap.

Surrogate data should be used cautiously. Use of surrogate data seems reasonable when the data utilized for the extrapolation are of high quality (i.e. if the extrapolation is made on ED_{50} s (potency), the ED_{50} s used must come from good dose-response curves that contain enough data points – two to four points are probably not enough). There must be an adequate understanding of the relationships among the data that are utilized for the estimation of the surrogate data.

For an endpoint like acetylcholinesterase enzyme (AChE) inhibition, a No-observed-adverse-effect-level (NOAEL) should not be used as part of an analysis like that presented in the current document. For AChE inhibitors, some data would be available indicating that AChE activity was depressed in at least one dose group. This information could be analyzed using some plausible dose-response model to get an estimate of potency, with perhaps a substantial margin of error. This, nonetheless, would probably be preferable to using some comparison of NOAELs for AChE activity.

It is imperative that the Agency be transparent in why and how they chose the surrogate data. Surrogate data should nearly always be associated with an explicit quantitative treatment of uncertainty so that the effects of that uncertainty on ultimate estimates of risk can be made. Not to use surrogate data in the kinds of cases mentioned could otherwise lead analysts to simply exclude chemicals that are expected to contribute some toxicity, thus biasing the analysis in an anti-protective direction. The arguments need to be stated clearly and convincingly. Entailed in such a process would be a discussion as to the "degree of confidence" or "level of uncertainty" the Agency has in the surrogate data.

Q2.2 How should situations be handled where an ED_{50} can be determined for many of the chemical members but cannot be determined for a few members? We would like the Panel's view on the use of NOAELs as substitutions for ED_{50}s for points of comparison.

The value of the data available to use for cumulative risk assessments is uncertain, in particular

when judged by individuals not seeing the entire set of information. Therefore, the quality of the data used in these processes should be closely scrutinized. Data collected years ago, by methods potentially not validated and/or current, could be inadequate for doing cumulative risk assessments.

Since NOAELs will be usually substantially lower than their corresponding ED_{50}s, it is potentially very misleading to mix them in the same analysis. It may be possible to combine both types of information by treating NOAELs as lower bounds on ED_{50}s using truncated-data methods. The required methodology would be similar to "survival analysis," where some patients have a known time of death and other patients are still alive, but all are included in the analysis.

Use of the NOAEL as a lower bound on the ED_{50} should not be used as a routine approach – the choice of action when development of an ED_{50} is not possible is really dependent on the data available. A number of situations are possible:

-The only data available are no-effect levels. If the highest tested dose were very small, the apparent potency of the chemical would be large if the NOAEL is used as the ED_{50}. This could conceivably lead to a situation where the cumulative risk assessment is driven by a chemical simply because it wasn't tested at a very high dose. Alternatively, the chemical may have been tested at several doses, including very high doses, without effect. Under this circumstance, the potency estimate derived from using the NOAEL as the ED_{50} surrogate would be low. There would be some confidence, based on the dose-response data available, that potency is in fact low and the chemical could be included in the cumulative risk assessment without introducing substantial error.

- There are effect data, but the data are regarded as insufficient to derive an ED_{50} . These effect data might be suitable for determining a lower bound estimate for the ED_{50} that would be better (i.e., closer to the true, undeterminable ED_{50}) than the NOAEL.

It was also suggested that a bounding RPF can be estimated based on a defined NOAEL of the index chemical. This point of comparison for the index chemical could be the NOAEL determined from the experimental study or a statistically determined level deemed as equivalent to no observable effect level (e.g., ED05, ED10). Accordingly, the accompanied uncertainties should be highlighted and be visible throughout the risk assessment.

Each of these possible situations leads to the conclusion that a decision to use the NOAEL (or some other value derived from the data) as a surrogate for the ED_{50} should be made on a case-by-case basis, with the overarching goal of selecting a value that minimizes the overall uncertainty in the cumulative risk assessment.

It should be noted that studies conducted under FIFRA to support a registration are required to show that the Maximum Tolerated Dose (MTD) has been reached. As such, this scenario would

imply that for this study, a LOAEL could be established for another endpoint, presumably more sensitive than the endpoint for which effects were not observed. Otherwise, the study would not have met the data requirement and would most likely have been replaced by a repeated study using a higher tested dose. Occasionally, a requirement for a repeated study is waved when a more sensitive endpoint (i.e., with lower NOAEL) has been identified in another study, perhaps in a different species or sex. Therefore, the fact that this scenario exists could indicate that further evaluation of the choice of endpoint or species or study may be necessary for establishing a different set of RPFs for cumulative risk assessment such that it reflects the use of a most sensitive endpoint.

Another scenario is when a "lowest-observed-adverse-effect-level" (LOAEL) can be established, but the particular dataset is insufficient for estimating the ED_{50} through model fitting. This could be due to having the lowest tested dose showing substantial level of effects, there is a substantial gap between the LOAEL and NOAEL, or, there is only one tested dose at which the endpoint for RPF is observed. When a LOAEL is available, it may be possible to establish an alternative RPF based on a comparison of effective dose at a comparable level of response. For example, if the LOAEL corresponds to a 35% of response, the RPF for this particular chemical could be based on the ED_{35}, instead of ED_{50}, of the index chemical (effective dose at 35% response). The RPF should be based on a dose level at a comparable level of response between the index chemical and the chemicals in the CAG. Moreover, the toxicity database should be viewed as a whole when selecting the endpoint(s) and datasets for establishing a set of RFP for cumulative risk assessment.

Question 3 We would like the Panel's view on the relative importance of the factors discussed in the paper for selecting an index compound.

The most important considerations in selecting an index compound are that 1) it acts toxicologically as purely as possible by the common mechanism defining the group (that is, there should be no other modes of appreciable toxicity in animals or humans), and 2) quantitative data for assessing potency be available for as many routes of exposure, genders, species and strains as possible that might contribute relative potency data to the analysis of the relative potencies of other chemicals. A cumulative assessment group (CAG) based on qualitatively similar pharmacokinetics (PK) would have better statistical behavior than a CAG that combines oxons with precursors that must be metabolized to their oxon forms. Stability of relative rankings across enzymes (e.g., plasma and red cell cholinesterase) is less important because the inhibition of these two enzymes should probably just be regarded as two different activities, each with its own separate toxicological implications.

In testing the assumption of parallelism as a component in evaluating the application of dose addition or in estimating relative potency using dose-response models that do not involve the assumption of parallelism, the selection of the index or reference chemical is totally incidental. To

assess parallelism, you must be able to estimate a slope. If the analysis is treated as a multiple regression (equation 9 or 15), the selection of the index chemical does not matter. The results of the analysis will be the same in terms of assessing the model fit. The relative potency estimates will of course vary with the selection of the index chemical but this does not impact the assessment of the model fit.

Since the RPF is a ratio at a pre-determined point of comparison, the choice of index compound would not be critical as long as a comparable point of comparison is used among all Cumulative assessment groups. However, the ultimate estimation of cumulative risk will be based on the point of departure of the index chemical, not its point of comparison. Therefore, it is important to model the RPF after a chemical that has the best and most complete data for the common endpoint(s). Having the best data for the index chemical could ensure that the uncertainties in the next step of risk assessment are minimized.

RPFs for three endpoints (i.e., plasma, RBC, and brain ChEI, and in males and females) are presented for male and female rats in the pilot study. If multiple endpoints will be used in characterizing the risk, selecting an index chemical with the best and most complete toxicity data would not necessarily mean using one index chemical for all three endpoints. Nevertheless, basing the RPFs for all endpoints on one index chemical would have an obvious advantage of allowing a direct comparison of the cumulative risks calculated from all endpoints.

Question 4 We would like the panel to comment on the log dose-probit analysis used to extrapolate the ED_{50}s for the chemicals evaluated in this pilot.

The Panel's discussion was hampered by confusion with a statistical technique called "probit analysis" which also uses the probit transformation to linearize a fractional response, but for counted data (e.g. the number of subjects out of *n* dying of toxicity at a given dose) . This methodology involves weighting for binomial error, which is not appropriate for the data presented for discussion. The pilot study used the probit transformation but did not use "probit analysis". To differentiate, the methodology used for the Pilot Study will be referred to as "probit transformation".

Probit analysis is appropriate for quantal dose response data, that is, data sets where the fraction of individuals experiencing a particular response is measured at a series of doses of the toxicant.[1],[2] The basic mechanistic presumptions of probit analysis are that 1) the effects occur when individuals' thresholds for the response are exceeded, and 2) the thresholds of different individuals

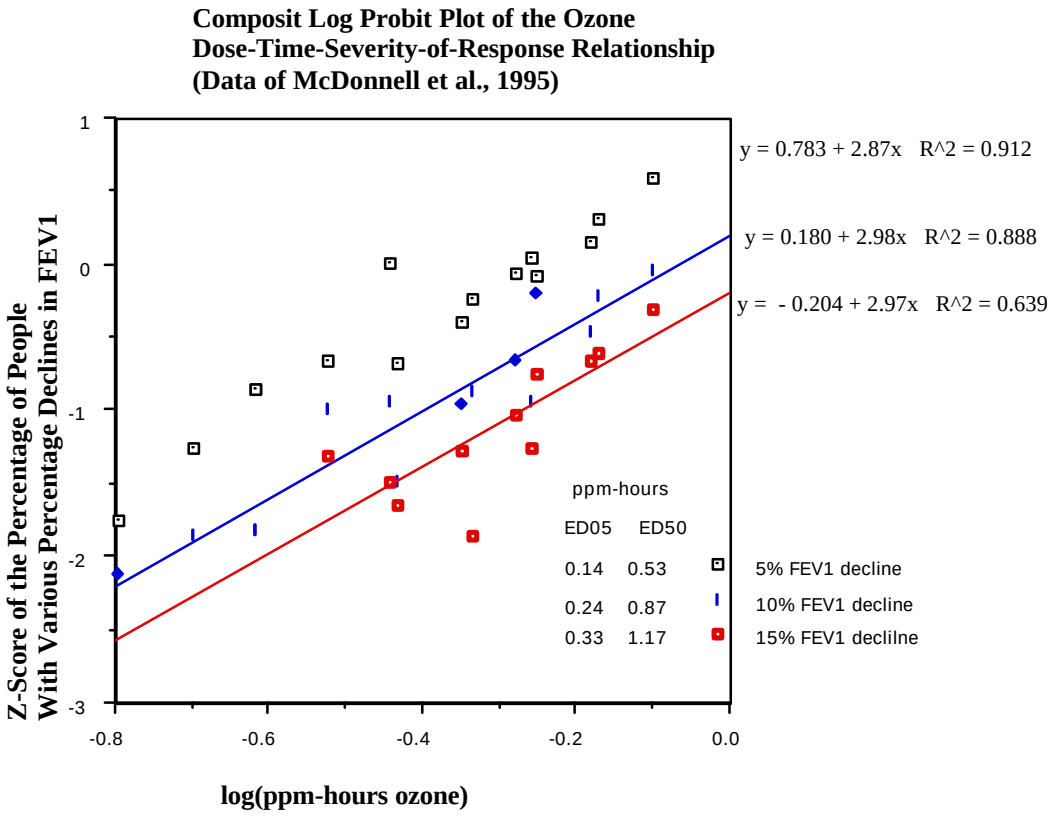
[1] Hattis, D., Banati, P., and Goble, R. "Distributions of Individual Susceptibility Among Humans for Toxic Effects--For What Fra
Does the Traditional 10-Fold Factor Provide How Much Protection?" Annals of the New York Academy of Sciences, Volume 89

[2] Hattis, D., and Silver, K., "Human Interindividual Variability--a Major Source of Uncertainty in Assessing Risks for Non-Cance
421-431, 1994.

have a lognormal distribution (the logarithms of the individual threshold doses have a normal Gaussian distribution). The lognormal distribution for individual thresholds is a natural expectation for cases where many factors contribute materially to individual differences, and those factors tend to act multiplicatively in affecting individual threshold doses.[3] The probit analysis framework can be used to analyze data for continuous outcome parameters (e.g. changes in FEV1 from ozone exposure),[4] by converting the dose response information to quantal form—e.g. by defining a series of "response" levels (e.g. 5%, 10%, and 15% change in FEV1) and counting the fraction of exposed individuals who respond at each level as a function of dose:

[3] Hattis, D. and Burmaster, D. E. "Assessment of Variability and Uncertainty Distributions for Practical Risk Analyses" Risk Ana

[4] Hattis, D. "Strategies For Assessing Human Variability In Susceptibility, And Using Variability To Infer Human Risks" In Hum Measures, Modeling, and Risk Assessment, D. A. Neumann and C. A. Kimmel, eds., CRC Press, Boca Raton, FL, pp. 27-57, 19



Data Source: McDonnell, W. F., Stewart, P.W., Andreoni, S., and Smith, M.V. 1995. Proportion of moderately exercising individuals responding to low-level, multi-hour ozone exposure.

The direct use of the probit transformation in an attempt to linearize cholinesterase inhibition data is inherently unusual and questionable. The implication is that individual molecules of the cholinesterase enzyme each have a threshold concentration of the pesticide, proportional to chronic external dose rate, at which they will become inhibited. It is not completely out of the question that, at least in the case of the brain acetylcholinesterase, even if the molecules are all identical, the molecules of the enzyme could be distributed in various locations in the brain that would put them at different risk of being inhibited. However, even in that case, unless one has a multimeric enzyme (like hemoglobin) where the affinity of one monomer affects the affinity of other components of the complex, one would expect a one-hit rather than a highly sigmoid or threshold-like response for individual acetylcholinesterase molecules as a function of concentration.

Some discussion of model selection should be given. The application of probit analysis to AChE inhibition, however, may not be the most appropriate approach. When applied to continuous response data, probit analysis does not have the same theoretical justification based on the

binomial distribution that probit analysis has for quantal responses. As discussed by Finney (1971, pp. 219 to 220), a logit transformation will often work as well or better. A serious analysis would involve the use of data on individual animals.

In a previous Panel review, the suggestion was made to use Michaelis-Menton kinetics. This does seem like a very natural approach in which AChE activity is constrained at zero but not constrained to any upper limit of activity (Vmax). The application of Michaelis-Menton kinetics would involve nonlinear modeling, which has its own set of problems in diagnostic tests for model fit. Linear transformations could be used but could be very awkward.

Another alternative would be to use a simple exponential model. This would probably give a reasonable fit to the data. This model fit the individual animal data reported by Guilhermino et al. (1998) very well. Conceptually, the exponential model is closely related to the one-hit model, a minor variant of the exponential model. When measuring AChE inhibition, the exponential model may be attractive because the interaction of the toxicant with the enzyme is only slowly if at all reversible. Thus, a one-hit sort of assumption for the dose-response model may be plausible in modeling the net effect on a population of enzymes.

The inhibition of individual molecules can be treated as a Poisson process. A cholinesterase molecule is either "hit" or not by a relevant inhibitor molecule, and the probability that it will remain active is just the probability of receiving zero Poisson "hits" or e^{-m}, where m is the number of "hits" per acetylcholinesterase molecule, proportional to the Area Under the Curve of Concentration X Time for the inhibitor. The average "hits" per molecule (m) can be found from percent (%) inhibition data from the following reasoning:

$$\text{Fraction ACHE Inhibition} = 1 - P_{0\,\text{hits}} = 1 - e^{-m}$$

$$e^{-m} = 1 - \text{Fraction ACHE Inhibition}$$

$$m = -\ln(1 - \text{Fraction ACHE Inhibition})$$

Therefore, a potency measure can be defined as this m "hits per molecule" per mg/kg-day dose rate at steady state.

There are quite a few potential sources of nonlinearity at high doses that could cause cholinesterase dose-inhibtion relationships to depart from the simple one-hit formula derived above. The high dose data points in particular often depart markedly from expectations of a simple one-hit saturating response. It is disappointing that the analysts in this case do not exhibit any of their actual data fits so that the reader can assess how well the data are actually described by their probit formula. Whatever equation or system of equations is used, there should be some attempt to assess whether the equation in fact is a reasonable summary description of the form of the data for enough individual data sets to make it convincing.

If the background document is intended to serve as a model for the assessment of cumulative risks, it is important for the analysis to be reasonably rigorous and transparent. In this respect, the decision to use group averages based on the data evaluation records (DERs) was inappropriate. The Agency has the ability to retrieve quickly full text copies of all of the studies submitted to EPA. Further, the studies will typically have tables that give the responses in individual animals. Using group averages, particularly averages for only the dosed groups that are expressed as a percentage of the control response, results in a substantial loss of information. The whole point of the document under review is to use a statistical analysis to reach or support a conclusion. Therefore, individual animal data should be used regardless of the dose-response model that is selected. This can be labor intensive. If it is too much of a burden for the case study of 24 chemicals, the number of chemicals covered in the case study could be reduced. For a case study, it is more important to do the analysis properly than to do the analysis for all compounds on which data are available.

In any statistical analysis that involves fitting a model to data (preferably data on individuals), statistical tests indicating that the model provides an acceptable fit to the data should be given. Such statistics are not presented in the current document. Thus, in a table such as Table 1a, it would be advisable to add a column giving the p-value corresponding to the F ratio for the model fit.

There is a need for weighting of the data points by their statistical strength. The current document makes no mention of this, but it should be vital to any good regression analysis in which the points differ greatly in their uncertainty. Minimally, if ordinary least squares regression analysis is used the points should be weighted with the inverse of their variances [(1/standard error of the mean)^2], where the standard error of the mean of the individual observations is traditionally calculated as the standard deviation divided by the square root of the number of observations.

The ED_{50} is a natural point of comparison for judging relative potencies because ED_{50}s in other cases (e.g. lethality responses) tend to be the most statistically robust for limited data sets. However, the point of this exercise is to predict the contributions to cholinesterase inhibition effects where each of several cholinesterase inhibitors is present at relatively low concentrations as residues in foods, water, etc. In this case, therefore, if there are high dose nonlinearities that become significant at the ED_{50} and somewhat lower, it might be advisable to assess the fit of whatever model is used to the available data points, and, in cases where the fit is poor, throw out the highest dose points and estimate a low dose potency (e.g., "hits" as defined above per unit dose or $1/ED_X$, where X could be 10 or 20% or some other fractional inhibition below 50.

Some attention should be given to how a default approach could be modified when data are available to support a more mechanistic approach. For example, with OP's, PBPK models for the individual compounds could be developed (some already exist) and then combined. Suitable experiments would be required to support the joining of the individual models. The combined

model could predict esterase inhibition for cumulative exposures without the need to worry about many of the assumptions that are part of the probit modeling approach. The overall guidance should clearly state the approach to cumulative risk assessment that will be used as a default, when the desired data and mechanism-based models (e.g., PBPK models) are not available, but also clearly indicate how the default approach would be modified as more data become available. This would serve to encourage interested parties to generate the data.

Greater use of uncertainty analysis should be made. Quantitative uncertainty analyses should be developed to the extent possible. This would help with the overall evaluation of the statistical modeling approach being used.

The Agency should use individual animal data, reconsider the selection of the probit model and, at very least, discuss why a particular model is selected and how well the model fits the data both conceptually and statistically.

Question 5 For this group of chemicals, OPP has sufficient data to calculate relative potency factors by the oral route for six different compartment/sex groups. Relative potency factors could be calculated by use of a single compartment/sex or by compiling data across compartment/sex groups.

Q5.1 If the Panel favors a single compartment/sex, please comment on the criteria that should guide the choice of a compartment/sex group.

The Panel generally felt that combining datasets across compartments and sex groups should be avoided. There is no compelling reason not to examine all three compartments in both sexes. This approach examines all the data that are available. Sex differences preclude the combining of males and females, and not all tissues are equally affected by all OPs, which argues against compiling data across compartments. While female rats tend to be more sensitive to some of the toxicants in this class, it is unclear that this would always be the case. Thus, selection of males or females as a rule may not be appropriate.

The Agency should not be too swift to distill the data down to a single set of RPF for risk assessment. First of all, the three compartments of ChEI do not represent the same endpoint of toxicity. While brain ChEI reflects the effects at the central nervous system, it has been the Agency's policy that blood ChEIs could be used as surrogate for the effects at the peripheral neural junctions for which data are largely not available. In this regard, it would not be appropriate to favor one ChEI endpoint over the others unless it is known that one endpoint is the most sensitive of the three for all chemicals within the cumulative assessment group. Unless the overall data could justify for combining the male and female data to an overall set of RPF, it would be advisable to use each set of RPFs from males and females separately, or to include a weighing factor for sex-specific sensitivity in deriving an overall RPFs for both sexes. The

subsequent risk characterization could be simplified from having one, instead of two sets of RPFs.

It has been argued that plasma cholinesterase has nothing to do with cholinergic transmission, and thus it should not be used as an endpoint for cholinesterase inhibitor toxicity. While plasma cholinesterase has nothing to do with neurotransmission, neither does RBC cholinesterase. The "missing" data therefore continues to be peripheral cholinesterase inhibition in many studies. Without these data, there should be continued support for using either blood enzyme as a surrogate for those measurements. One the other hand, another Panel member argued that plasma AChE values should not be used and that the Agency should avoid use of plasma AChE values. Plasma AChE activity may be a good assay of exposure but it is not clear that it is a meaningful toxicologic response.

> **Q5.2** It is proposed in this pilot analysis, that data could be compiled across different studies to provide more confidence in the determination of relative potency. In establishing an effective dose (e.g., ED_{50}), to what extent should one compile data for each chemical of interest within or across different measures and/or studies? What are important criteria to consider when compiling data?

The pitfalls in attempting to combine the results of studies from different times, places and investigators are addressed by what is known in the clinical sciences as "meta-analysis". A good general reference is *H Cooper & LV Hedges (editors), The Handbook of Research Synthesis, New York: Russell Sage Foundation, 1994.*

In combining different studies, only studies of acceptable quality would be used. Studies that are obvious "outliers" should not be used. However, there needs to be a transparent explanation as to why a given study was deleted from the compilation of studies. As a first step, the response/dose/time data for the individual animals should be fit to some appropriate dose-response model such as:

$$Y = \beta_x f(x) + \beta_t g(t) + \alpha$$

where β_x is the slope for dose, β_t is the slope for time, the functions f() and g() are transforms on dose and time, and Y is the response of a transform of the response. The expectation would be that the model would fit the studies conducted on the same compound. If the model did not fit the data, the assumption of a common slope for the same chemical would be contradicted. Occasionally, this would happen for random or at least unidentifiable reasons. The same sort of thing may happen in comparisons of slopes among chemicals (Question 7). Nonetheless, the frequent failure of the model to fit the data would suggest that whatever procedure is being used in the analysis may be flawed. In other words, if the procedure being used does not generally indicate that Chemical A is Chemical A, the procedure may not be very useful for determining whether or not Chemical A through Chemical Z appear to act by a common mechanism. Thus, the combination of data on a single chemical might not only lead to better estimates of potency

but could also lead to an evaluation of the analytical method used to test the assumption of dose addition.

Regardless of what endpoint(s) to be used, the Agency is commended for developing such a sizable database. Besides being useful for eventually establishing RPFs for addressing cumulative risk of organophosphates, it appears that this database could also be useful for furthering the understanding of various aspects of the toxicity of organophosphate pesticides. For example, the database might be useful for evaluating the traditionally applied default uncertainty factor of 10 to extrapolate the toxicity threshold of these endpoints from intermediate- to long- term. Data presented in this pilot study appears to indicate that a steady state is reached at or around 3 months for ChEI endpoints. With this database, further investigation into the need for an uncertainty factor for subchronic-to-chronic extrapolation may now be possible.

Question 6 Dose addition is considered an appropriate default approach to cumulative risk assessment. The mathematical definition of dose-addition requires a constant proportionality between the effectiveness of the chemicals being considered. It is anticipated that extensive dose-response data will not be available for many chemicals. Please comment on the approach taken to evaluate parallel dose-response curves. Please comment on how rigorous an analysis is needed to evaluate the assumption of parallel dose-response curves.

The document indicates that two tests of parallelism were used: an F-test and a Spearman rank order correlation of the log(ED_{50}) values and the intercepts (p. 16, last paragraph). It is not clear what was done with the F-test. The statement *"In other words, $p<0.05$ for the F-test would indicate the slopes of the linear regressions were statistically different."* is confusing.

The second test for parallelism, the Spearman rank order correlation of the log(ED_{50}) values and the intercepts, seems to be clearly inappropriate, particularly if dose is on a log scale. The document states that "*Based on the assumption of parallel dose-response curves, the relative rank of the calculated log(ED50)s should be exactly the same as the relative order of the intercepts.*" (p. 10, last paragraph). This is incorrect if the statement is referring to the y-intercept. Under the assumption of parallelism, the rank order of the y-intercepts should be the reverse of the rank order of the log(ED_{50}) or the log of any other effective dose. This is visually illustrated in the top part of Figure 1 on page 17 of the document under review.

Regardless of the sign of the correlation, however, a rank order correlation is not a meaningful statistical test for parallelism. The slopes estimated from the individual studies will have variability and lines will cross at some point. Even on a linear scale, the propensity to cross over will depend both on the differences in the slopes as well as distance of the ED_{50} values from the intercept (i.e., a dose of zero). When dose is expressed on a log scale, the intercept (a dose of 1 in some unit corresponding to a log dose of zero) will change depending on how the dose is

scaled. The relationship is simple:

$$\alpha_{new} = \alpha_{old} - \beta \log(F)$$

where *F* is the scaling factor. Thus, if the slope is 2.5 and the intercept is 2 on a log dose scale expressed in units of mg/kg, the intercept will be -5.5 using a dose scaling (*F*) of 1000 to convert to μg/kg.

An example of the problem in using rank order correlation can be seen in the bottom part of Figure 1 in the Agency's background document. Taking the dose units in the figure as mg/kg/day, the Spearman rank order correlation is about -0.94. If the doses are scaled to μg/kg, the Spearman rank order correlation is -1.0. If the doses are scaled to g/kg, the Spearman rank order correlation is -0.4. Thus, the results of a rank order correlation are dependent on how the dose is scaled. This is not a desirable characteristic in an evaluation of parallelism and it does not test the underlying model (equation 9). Different models, like logit transformations and weighted regression could be used, instead.

The interpretation and reliability of the intercept changes with the potency of the drug, while the reliability of the slope and the ED50 depends on the experimental design.

The underlying experiments were not designed to address the question of parallelism. Relatively little weight should be given to how frequently some *ad hoc* statistical analysis of slopes (for a model that is dubious in this application in the first place) has sufficient power to detect differences at some chosen P level. The basic case for grouping anticholinesterase agents together is knowledge that they inhibit the same functionally important enzymes in roughly the same places in the body. If empirical support is deemed helpful, then the way to proceed is to imagine some likely mechanistic sources of changing quantitative inhibition behavior with dose, and deliberately design experiments to detect these and determine at what percent inhibition levels they become large and troublesome enough to warrant excluding otherwise relevant data. More complex, biologically motivated approaches, such as PBPK modeling, would not require some of the constraining assumptions inherent in the probit approach, such as parallelism of dose-response curves.

To identify likely sources of high dose departures from low dose patterns, we first need to make the quantitative case for what the usual pattern of change of cholinesterase levels should be as a function of dose in a well-behaved equilibrium system. At the heart of the argument is simple bimolecular reaction kinetics. If we have a concentration of cholinesterase molecules floating free in a well mixed pool, such as plasma, then the rate of the irreversible covalent reaction between the organophosphate and the relevant serine residue on the cholinesterase will depend linearly on the concentration of each reactant:

Rate of Inactivation Reaction = k[Inhibitor][Cholinesterase]

In a chronic continuous dosing experiment, this reaction will not be the only relevant factor determining cholinesterase levels. Even in the absence of inhibitor, there will be some rate of loss of the cholinesterase enzyme molecules because of the action of various proteases, some rate of natural denaturation, and other processes. In general, these losses must be balanced by ongoing new synthesis of cholinesterase molecules to maintain the baseline level of cholinesterase observed without any dosing with cholinesterase inhibitor molecules. Imagining constitutive synthesis, and first-order Natural Loss of cholinesterase molecules with a rate NL, the baseline condition must be:

Constitutive synthesis rate = NL[Cholinesterase baseline equilibrium]

Combining these equations, after chronic low dose administration of the inhibitor, there will be a new balance with a new equilibrium level of cholinesterase

Constitutive synthesis rate = [Cholinesterase](k[Inhibitor] + NL)*

Solving for the new equilibrium level of cholinesterase yields

$$\text{Cholinesterase new equil level} = \frac{\text{Cholinesterase synthesis rate}}{k[\text{Inhibitor}] + NL}$$

$$\text{Cholinesterase new equil level} = \frac{NL[\text{Cholinesterase baseline equilibrium}]}{k[\text{Inhibitor}] + NL}$$

The ratio of the "new equilibrium level" to the "baseline equilibrium" is the fraction of the enzyme that escapes inhibition. This reasoning therefore leads us to conclude that if everything behaves as specified, then 1/fraction cholinesterase uninhibited is linearly related to the concentration of inhibitor (k[Inhibitor] + NL) which in turn should be linearly related to external dose rate at least at low doses.

At high doses, the following could happen:

There could be nonlinearities in the internal concentration of inhibitor in relation to external dose rate. This can happen, for example, if under low dose conditions the inhibitor reacts with another depletable molecule, or if it is destroyed by a saturable detoxification system. It could also happen that by depleting free uninhibited cholinesterase itself, a mode of loss of the inhibitor is reduced, leading to higher steady state inhibitor concentrations (or AUC) per unit external dose.

The inhibitor could interfere with the constitutive cholinesterase synthesis, causing greater

inhibition than would be expected at high doses.

There could be a homeostatic control system that senses inadequate cholinesterase levels and causes enhanced synthesis of cholinesterase molecules that reduces the degree of inhibition that would be expected under the simple constitutive synthesis case.

As mentioned earlier, effectively k could change with dose, at least in the brain, if some acetylcholinesterase molecules are in locations that make them more or less susceptible to inhibition as a function of inhibitor concentration in the blood. As the dose rate increases, the cholinesterase molecules in the more susceptible locations are preferentially inhibited, leaving molecules in more protected locations with (effectively) lower average "k" values in relation to external dose or serum concentrations.

Each of these possibilities (and possibly some others) can be directly assessed via specific experiments with model compounds.

Question 7 How does one handle a response for a chemical that displays a different slope (i.e. an outlier)? Examples were demonstrated in this pilot analysis where one or a few chemicals of the common mechanism group exhibited pronounced species, sex, or compartment differences from the majority of chemicals.

When using the probit approach, or a similar statistically constrained approach, uncertainty analyses, preferably quantitative uncertainty analyses, should be used to evaluate the consequences for the analysis of alternative approaches to the handling of outliers. For example, perhaps the outlier data are replaced with surrogate data for a compound that is not an outlier and whose potency is at least as great as the potency of the outlier. The effect on the overall assessment of using various surrogates could be examined, with an accompanying consideration of how uncertainty is affected by use of the surrogate data. Alternatively, approaches that are more mechanism based, such a PBPK/PD modeling, would remove the need for constraining assumptions such as parallelism of dose-response curves. When uncertainty of the statistically constrained approach is high, an incentive for the use of the more mechanism-based approach exists.

If a true outlier is found using an appropriate dose-response model, some judgment will be necessary. An internal validation of the analytical method used to test the assumption of dose addition could involve combining data sets for the same compound. This would lead to a better appreciation of how often an apparent difference in slope might be attributable to variability in the available data and/or limitations in the analytical method used to assess parallelism.

Here it is worth noting that the assumption of parallelism is based on the further assumption that the animals used in the bioassays are randomly taken from the same population. Pesticide

bioassays on file with the Agency have been conducted at different times, in different locations and may differ in other ways that are not readily apparent. Thus, the failure to demonstrate parallelism may have a more complex interpretation than simply suggesting that the mechanisms of action are different.

Substantial reliance should be placed on what we know about the mechanism of action. If the slope for a compound is substantially and significantly different from other slopes in a presumed class of compounds and if parallelism in the dose-response curves is required by the model being used in the analysis, the determination of whether or not to include the compound in the assessment of cumulative risk for the class could be guided largely by the confidence in the knowledge that the mechanisms of action for the chemicals in the presumed class are properly understood.

Perhaps most importantly, the issue of outliers based on a failure to estimate parallel slopes may be incidental and irrelevant if the analysis uses an inappropriate dose-response model.

REFERENCES

Finney DJ. 1971. Probit Analysis. 3rd ed. Cambridge University Press, Cambridge, UK.

Guilhermino L; Soares AM; Carvalho AP; Lopes MC. 1998. Correlation between whole blood cholinesterase activity and cerebral cortex cholinesterase activity in rats treated with parathion. Chemosphere. 37(7):1385-93.

APPENDIX

The following are comments from one Panel member, Dr. Pat Durkin that are not specifically a response to a particular question posed to the Panel. However, the comments are relevant and are included in this appendix to the Report.

General Assumptions

The document is based on three 'assumptions' (detailed on p. 10): a common mechanism of toxicity, the absence of interactions, and a 'constant proportionality among the effectiveness of the chemicals'. The analysis presented in the document, however, focuses only on the third assumption in which an attempt is made to determine if a the log-dose probit-response model can be used to meaningfully characterize relative potency. I think that each of the assumptions must be more carefully examined.

Common Mechanism **-** I do not think that there will be any substantial debate about the first assumption: a common mechanism of action in the inhibition of AChE activity. There may be some concern about the other endpoints that one or more of the subject chemicals may cause but I consider this to be beyond the scope of the current analysis and review.

While the common mechanism of action may be widely accepted based on what we know of these compounds and may not require extensive review, I think that the document would be enhanced if a reference was made to the recent OPP document, *The Use of Data on Cholinesterase Inhibition for Risk Assessments of Organophosphorous and Carbamate Pesticides* (draft dated August 18, 2000). I think that more emphasis on mechanistic information could be important for this class of compounds as well as other classes, particularly when other types of information (such as an analysis of dose-response relationships) appear to contradict the assumption of dose addition. As I discuss further in my response to Question 7, it may be reasonable at times to use mechanistic information to support the assumption of dose addition when this assumption appears to be contradicted by dose-response data. Some quantitative approaches to assessing interactive classes have been proposed (Durkin et al. 1995) and might sometimes be useful. A good qualitative discussion of the mechanistic data, however, would probably be sufficient for the current document.

The discussion or understanding of the mechanistic data is impeded by the use of codes in the document. I am not sure why this was done and I note that it was also done in the guidance on cumulative risk assessment. I do appreciate that certain types of FIFRA data are considered proprietary and cannot be divulged to the public but I believe that the identity of the active ingredient as well as the results of toxicity tests on the active ingredient can be released (U.S. EPA 1985). The Agency may want to consider adding a key specifying the names of the chemicals as an appendix. This would help in a fuller discussion of the data which supports the assumption of a common mechanism of action.

Absence of Interactions **-** The second assumption, the absence of interactions, is very important and deserves more attention. The issue of cumulative risk is essentially an issue in the risk assessment of mixtures. While the document addresses this somewhat indirectly with the discussion of dose addition and relative potency, the document should reflect the existing Agency guidelines on mixtures' risk assessment (U.S. EPA 1987) as well as the draft guidance for the risk assessment of mixtures (U.S. EPA 1999). The Guidance on Cumulative Risk Assessment (June 22, 2000, p. 87) does mention these EPA mixture documents and has a very short paragraph on assessing the potential for interactions (p. 28). I do not find

a discussion of the potential for toxicologic interactions in *The Use of Data on Cholinesterase Inhibition for Risk Assessments of Organophosphorous and Carbamate Pesticides* (August 18, 2000).

The one paragraph discussion of interactions on page 10 of the current document is cursory. I am not intimately familiar with the interaction studies on OPs but a quick look at the titles in the reference list of the document suggests that synergism has been reported in at least some studies. In addition, there is at least one possibly relevant study that is not included in the literature cited - i.e., Su et al. 1971. If the document under review is to be the basis for a "model approach", there should be at least a subsection that describes these studies in more detail and addresses the weight-of-evidence for the assertion of dose addition relative to some form of interactive joint action.

The last sentence in the paragraph on interactions states: *Thus, the assumption of additivity appears reasonable, particularly at lower exposures.* As I note above, I do not think that this conclusion is supported by the analysis (or lack of analysis) presented in the document. In addition, there is an implicit assumption that we are only worried about "low exposures". The definition of 'low exposure' is not easy. Lastly, if the approach outlined in the document is intended to serve as a model for the assessment of cumulative risk, it should consider both low and high exposure levels. Again, the recent guidance from the U.S. EPA (1999) on the risk assessment of mixtures does this.

Dose Addition, Potency, and Parallelism - For the third assumption, the wording of the start of this section suggests the need for some clarification:

> *There is a constant proportionality among the effectiveness of the chemicals. In other words the dose-response curves of the chemical group were parallel.* (p. 10)

The first sentence is correct but the second sentence is only correct for some classes of dose-response models such as the probit model. The document also states that:

> *The assumption of parallel dose-response curves is a major principle in cumulative assessment.* (p. 56)

This again is correct only in the application of some dose-response models. Because the document is so focused on tests for parallelism in an apparent attempt to justify the assumption of dose addition, I think that a more precise summary of the relationship of parallelism to dose addition is necessary.

The basic concept of dose addition requires a dose-independent ratio of equitoxic doses. Adopting the nomenclature of Finney (1971), this ratio is the relative potency (ρ_i) of the i^{th} chemical with respect to some reference chemical and is defined as the ratio of an equitoxic dose of the reference chemical, ζ_r, divided by an equitoxic dose of the i^{th} chemical, ζ_i,

$$\rho_i = \frac{\zeta_r}{\zeta_i} \qquad \textbf{(3)}$$

where ζ can represent some relevant measure of toxicity such as the ED_{50} or ED_{95}. The essence of dose addition is the ability to add the dose of the reference chemical (x_r) and the potency (ρ_i) weighted doses (x_i)

of other chemicals in a mixture to obtain an equivalent dose of the reference chemical:

$$x_{r_{equiv}} = x_r + \sum_{i=2}^{n} \rho_i \cdot x_i \qquad \textbf{(4)}$$

A constant relative potency is thus required in any application of dose addition.

The second sentence, the requirement for parallel dose-response curves, is dependent on the dose-response model being used and is not a general requirement of dose addition.

In the current document, EPA has used the log-dose probit-response model. This model and other similar models such as the logit transformation form the basis for much of the classical literature on dose addition (e.g., the work of Finney and Bliss). The model has the general form:

$$Y = \alpha + \beta \log(x) \qquad \textbf{(5)}$$

where x is the dose, β is the slope of the dose-response curve, is the intercept, and ***Y*** is some measure of response or a transform of the response such as a probit or logit. The logit and probit models are often used in the evaluation of dose addition because much of the data on the joint action of toxicants consists of acute studies involving quantal responses where either the probit or logit models provide adequate fit to the available dose-response relationships.

Under any dose-response model in the form of equation 3 (e.g., the probit or logit), parallel slopes are required. For any two chemicals, equitoxic doses (ζ_1 and ζ_2) are, by definition, given as:

$$\alpha_1 + \beta_1 \log(\zeta_1) = \alpha_2 + \beta_2 \log(\zeta_2) \qquad \textbf{(6)}$$

If the slope of the dose-response curves are identical to some value, β, ($\beta = \beta_1 = \beta_2$), then

$$\zeta_1 = \left[e^{\frac{\alpha_2 - \alpha_1}{\beta}} \right] \cdot \zeta_2 \qquad \textbf{(7)}$$

and the relative potency of chemical 2 with respect to chemical 1 (ρ_2) is a constant independent of the value of x_1 or x_2:

$$\rho_2 = \frac{\zeta_1}{\zeta_2} = e^{\frac{\alpha_2 - \alpha_1}{\beta}} \qquad \textbf{(8)}$$

Thus, any dose of chemical 2 can be converted to an equivalent dose of chemical 1 by a constant:

$$x_1 = \rho_2 x_2 \qquad \textbf{(9)}$$

as required by equation 2.

Note further that if the slopes in equation 4 are not equal ($\beta_1 \neq \beta_2$), equation 5 becomes

$$\zeta_1 = \left[e^{\frac{\alpha_2 - \alpha_1}{\beta_1}} \right] \cdot \zeta_2^{\beta_2} \qquad \textbf{(10)}$$

and a potency is not constant across all doses of chemical 2. This is why parallel dose-response curves are required for dose addition in the probit and logit models. Generalized for ***n*** number of chemicals, the log-dose probit-response model is:

$$Y = \alpha_r + \beta \log(x_r + \sum_{i=2}^{n} \rho_i \cdot x_i) \qquad \textbf{(11)}$$

where $_r$ is the intercept for the reference chemical, β is the common slope for the dose-response curves of the individual chemicals in the mixture, and ρ_i is the relative potency of the i^{th} chemical in the mixture relative to the reference chemical.

Parallelism, however, is not required in all dose-response models and one example of this is the simple exponential model:

$$Y = Y_0 e^{\beta x} \qquad \textbf{(12)}$$

where ***Y*** is the response, Y_0 is a parameter estimate of the control (zero dose) response, ***x*** is the dose, and β is a potency/slope parameter. In this type of model, parallelism of the dose-response curves is not a meaningful concept even though the definition of relative potency (equation 1) is identical to the definition under probit or logit analysis.

In the exponential model, relative potency can be calculated following the same general approach taken for the probit analysis example. The only special constraint is that the control responses (Y_0) are equal or can be normalized to be equal:

$$Y_0 e^{\beta_1 \zeta_1} = Y_0 e^{\beta_2 \zeta_2} \qquad \textbf{(13)}$$

Dividing both sides of the equation by Y_0 and then taking the natural logarithm of both sides,

$$\begin{aligned} \beta_1 \zeta_1 &= \beta_2 \zeta_2 \\ \zeta_1 &= \frac{\beta_2}{\beta_1} \zeta_2 = \rho \zeta_2 \\ \textit{where } \rho &= \frac{\beta_2}{\beta_1} = \frac{\zeta_1}{\zeta_2} \end{aligned} \qquad \textbf{(14)}$$

which is identical to the definition of potency under the probit model (equation 1) where chemical 1 is the reference chemical and chemical 2 is the chemical whose relative potency is being estimated. Thus, for the exponential model, relative potency is as the ratio of the slopes of the i^{th} chemical divided by the

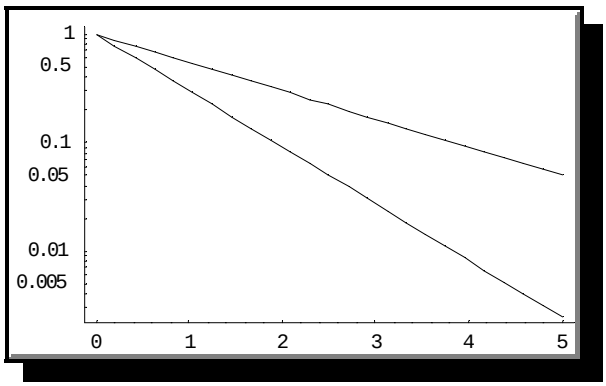


Figure 1: Illustration of linearized exponential model for two hypothetical compounds (Y_0=1, β_1=-0.6, β_1=-1.2).

slope of the reference chemical.

Again, the parallelism of slopes is not a meaning concept for this dose-response model. Using the linear transformation of the exponential model, which would probably be used in any analysis of real data:

$$\log_e(Y) = \log_e(Y_0) + \beta x \quad \textbf{(15)}$$

comparative plots of different chemicals would lead to a series of lines radiating from a common y-intercept, as illustrated in Figure 1. No common transformations of the exponential model lead to parallel dose-response curves and any transformation of the exponential model that did lead to parallel dose-response curves would have nothing to do with validating the assumption of dose addition.

When generalized for ***n*** number of chemicals, the exponential model takes the following form analogous to equation 9 for probit analysis:

$$\log_e(Y) = \log_e(Y_0) + \beta_r \left(x_r + \sum_{i=2}^{n} \rho_i \cdot x_i\right) \quad \textbf{(16)}$$

Given the definition of relative potency (β_i/β_r), this is equivalent to:

$$\log_e(Y) = \log_e(Y_0) + \beta_r x_r + \sum_{i=2}^{n} \beta_i \cdot x_i \quad \textbf{(17)}$$

In other words, under models like the exponential model, parallelism is not a requirement and the calculation of relative potency is unnecessary. All that the concept of dose addition requires from a model such as this is that the data all fit a common dose-response function.

The practical importance of this fact to the current document deserves emphasis: if you have a group of chemicals that are presumed or known to act by the same mechanism but these chemicals have apparently different slopes under a dose-response model that requires parallel slopes for dose addition, you may be using the wrong dose-response model. In other words, the lack of parallelism is not necessarily an indicator that the assumption of dose addition is contradicted. In most cases, the use of mechanistic data will be a better approach to assessing the assumption of dose addition than an empirical analysis of dose-response curves.
innapropriate

REFERENCES

Durkin PR; Hertzberg RC; Stiteler WS; Mumtaz M. 1995. The Identification and Testing of Interaction Patterns. Toxicology Letters, 79:251-264.

Goldstein A; Aronow L; Kaman SM. 1974. Principles of Drug Action: The Basis of Pharmacology. 2nd ed. John Wiley and Sons, New York, NY. 854 p.

Mumtaz MM; DeRosa C; Durkin PR. 1994. Approaches and challenges in risk assessments of chemical mixtures. In: Toxicology of Chemical Mixtures. Yang R (Ed). Academic Press, Chicago, IL. pp. 565-597.

Su MQ; Kinoshita FK; Frawley JP; DuBois KP. 1971. Comparative inhibition of aliesterases and cholinesterase in rats fed eighteen organophosphorus insecticides. Tox. Appl. Pharm. 20(2):241-9.

U.S. EPA (U.S. Environmental Protection Agency). 1985. Disclosure of Reviews of Pesticide Test Data. Feg. Regist. 50(229): 48833-48835. Nov. 27, 1985.

U.S. EPA (U.S. Environmental Protection Agency). 1987. The Risk Assessment Guidelines of 1986. Prepared by the Office of Health and Environmental Assessment, Office of Research and Development, Washington, DC. EPA/600/8-87/045. Available from NTIS, PB-88-123997/AS.

U.S. EPA (U.S. Environmental Protection Agency). 1999. Guidance for Conducting Health Risk Assessment of Chemical Mixtures. NCEA-C-0148, April 1999, External Scientific Peer Review Draft.