HOUSATONIC ECOLOGICAL RISK ASSESSMENT: PEER REVIEW EVALUATION

1. WAS THE ECOSYSTEM OF THE HOUSATONIC RIVER WATERSHED PROPERLY CHARACTERIZED, AND WAS THIS INFORMATION APPROPRIATELY APPLIED IN THE PROBLEM FORMULATION AND SUBSEQUENTLY IN THE ERA?

There is no doubt that much effort and thought have been undertaken by the U.S. EPA and GE. I applaud them and congratulate them on the tremendous amount of research, thought and effort that went into this ERA. It is always easy to "Monday-morning quarterback" and criticize the resulting documents. I hope my criticisms are not taken in this manner. They are meant to suggest my concerns and perhaps point towards some "confirmatory" sampling that may lessen some of the uncertainty associated with various assessment or measurement endpoints.

The Housatonic River ecosystem ("Ecosystem") was very well characterized. The report by Woodlot (2003) provides an excellent overview of the variety of studies conducted and species sampled or identified. However, it is unfortunate that neither bats nor waterfowl (examples: mergansers, fish-eating ducks or "dabbling" ducks that filter invertebrates from sediments) were included in the Risk Assessment Conceptual model (diagrammed in Fig 2.7-1 or included as assessment or measurement endpoints (Table 2.8-1). Also, given the importance to this ERA in assessing the risk to terrestrial species, including amphibians and salamanders, that less effort was devoted to sampling soil invertebrates such as earthworms, isopods, beetle larvae, millipedes and centipedes. Given that these were sampled in 2000 (Woodlot Report), and given the numbers of isopods, cicadas and slugs sampled, these may well be important routes of exposure for predators.

Finally, I am concerned that there is no comprehensive overview or integration of the species in the ecosystem mapped out in Figure 2.7-1. Given the conclusions (see more on this below) of "high risk" to benthic invertebrates, that information is not carried into the risk assessment for insectivorous birds. The birds are judged on strictly other, limited information. However, if their food base is judged to be at high risk, then the predatory organisms relying on this food base must also be judged to be at risk. In any case, there appears to have been little-to-no overview linking the variety of seemingly-independent assessment endpoints. After all, this is an <u>ecological</u> risk assessment – and one must view the resident species as an interacting biological community.

These ideas were discussed in another context during the Lenox public meeting. In my opinion, the EPA should review the prior documents detailing the goals of this ERA. An earlier memo states that the goal is to "ensure recovery and maintenance of healthy local populations and communities of biota." The question is: "how was this defined for the purposes of the risk Housatonic Rest of River assessment?

2. WAS THE SCREENING OF CONTAMINANTS OF POTENTIAL CONCERN (CPOCs), SELECTION OF ASSESSMENT AND MEASUREMENT ENDPOINTS, AND THE STUDY DESIGNS FOR THESE ENDPOINTS APPROPRIATE UNDER THE EVALUATION CRITERIA?

The screening of CPOCs was extensive and appropriate. The use of Ingersoll and Long's approaches for estimating threshold effect levels has been shown to result in conservative estimates of sediment-bound contaminant effect levels. I will have more comments, below, on the use of "hazard quotients" in the ERA process. Such quotients are a beginning and should be used only when the variances in denominator and numerator are understood or estimated. Without an estimate of the variances "hidden" by the derived variable, there can be little confidence in assessing degrees of risk in numbers above an HQ of 1. In essence, one could ask "is there any difference in an HQ of 1, relative to an HQ of 5, 7 or even 10?" How can one tell, without knowing the coefficients of variation in numerator and denominator? I would claim that, without this information, one cannot tell. This point was also discussed during the Lenox meeting. There should be greater emphasis on the use of species sensitivity distributions (or exposure distributions), much like has been used for fish (see below). An excellent reference is Posthuma, *et alia* (2002).

To this end, Appendix C.5, concerning point estimates and UCL calculations is excellent and provides an excellent rationale for, and approach to, the use of UCLs in this ERA.

As to the selection of assessment and measurement endpoints, I will have more to say on each one, below. However, one glaring need in this ERA is to include bats. They are present during breeding (Woodlot Report, Attachment C) and are very often prolific insectivores – and often focus on aquatic insects. Bat houses are relatively easy to construct to attract bats (much like the houses for swallows and other avian species) and would have allowed a good estimation of COPC accumulation and effects. I would also liked to have seen more terrestrial invertebrate analysis.

One of the comments brought up during the Public Meeting questioned the extent of sampling downstream from the PSA. It is my understanding that relatively few samples have been collected (n = 66). If this is so (I could not confirm by studying the Corps report (Final Supplemental Information Work Plan) or the Weston July 2003 report), then there should be continued efforts towards understanding the distribution of total PCBs downstream from Reach 9.

3. FOR EACH OF THE EIGHT ASSESSMENT ENDPOINTS EVALUATED IN THE ERA (LISTED IN ATTACHMENT B), ADDRESS THE FOLLOWING QUESTIONS:

<u>Assessment Endpoints for the Housatonic "Rest of River" ERA:</u>

3.1 Survival, Growth, Reproduction And Structure Of The Benthic Invertebrate Community.

1.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

By-and-large, yes.

1.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

Yes.

1.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

By-and-large, yes. However, there is still uncertainty added to the results by mixing sediments collected at one site into one "batch." In a synoptic approach, it would be better to determine the variance in PCB concentration from distinct samples and exposing laboratory organisms to the various field-assessed range of concentrations.

1.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes, particularly the toxicity endpoints using surrogate species.

1.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

The multidimensional scaling (MDS) approach is similar to a principal components analysis (PCA), as acknowledged in the text (pg. D-67). I would suggest that a log-linear categorical approach (Sokal & Rohlf 1989) also be employed, as the categorical approach to percentage responses (not only toxicity endpoints, but dominant species groups abundances) and PCB concentrations may yield further supportive information concerning community-level responses to PCB exposure. Finally, the data collected are most appropriate for canonical correlation techniques, such as detailed in Tabachnick and Fidell (1996)

The use of Shannon-Wiener diversity, H, to describe the benthic invertebrate community must be questioned. There are numerous problems with it: if the species richness is "low" (fewer than 100 species), then H' is a insensitive measure of the relative frequency of species (May 1975). Further criticisms of the use of H' include (Green 1979; Moriarty 1999; pg 242): H' is unaffected if one species replaces another (there is no consideration of the differences in species taxonomy among sites); communities are not "supra-organisms" and should not have a biological meaning to it; any index needs taxonomic identification to species level – not sub-family. The data are present to completely and successfully analyze total abundances, relative frequencies (perhaps employing a dominance index), or discuss a variety of indices, including the index of biotic integrity, IBI.

1.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

The degree of risk to aquatic macroinvertebrates is judged to be high, and I strongly concur.

1.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

The only suggestion I would make on the analysis of toxicity endpoints here (see Figures D.3-12 through D.3-17) would be to use an approach increasingly used in comparative toxicity (SSDs, in Posthuma, *et al*., 2002) in which the cumulative percentage of endpoints responding are plotted against their respective concentrations inducing the response. As an example, one could re-cast Figure D.3-12 as "Cumulative percentage of endpoints responding to a given total PCB concentration." It presents a more quantitative approach, rather than having the "overview" of drawing a conclusion from trends. In the SSD approach, one could categorically state (for Figure D.3-12), "75 percent [my example estimate; not a calculation!] of the endpoints responded at a tPCB concentration of 10 mg/kg or less" and "30% of the endpoints responded at 1 mg/kg or less." The value of this suggested "cumulative response" approach is in having a <u>quantitative</u> statement of what percentage of endpoints are expected to respond at a given PCB concentration. As the results are now presented, there is too much latitude (in my opinion) in how one could assess endpoint response.

Chronic endpoints should be separated from acute endpoints and given more weight. This should hold for the fish, amphibian, mammal and avian assessments, as well. To best understand the nature of the benthic responses, the analyses should use the synoptic chemistry data, as much as is possible.

1.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

The weight of evidence approach for these data was appropriate.

1.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Yes.

1.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

Yes.

2. Reproductive Success, Development, Maturation, And Condition Of The Amphibian Community.

2.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

Yes, there was a good attempt to gather information on the exposure of the two ranid frog species. The laboratory toxicity studies, with associated larval endpoints, provide particularly strong evidence. The EPA leopard frog field study unfortunately led to small sample sizes (this may be indicative of long-term population trends in the Housatonic River Valley). However, as per page 4-32 of the 4.4.1.1.1 section, using "visual interpretation" to understand results, particularly with small data sets, can lead to dramatic errors in interpretation – if only because of the sampling variance inherent with small sample sizes.

2.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

The field surveys were conducted and follow accepted scientific practices. However, the duration of the GE wood frog field study was relatively short and, given the limited numbers of egg masses and limited numbers of ponds in the contaminant distribution classes. This is not, *per se*, incorrect – but the study has low power. The use of the field tests, with the associated variances in distribution and body burdens, should not be used to mask the more direct results from laboratory toxicity studies by Fort, *et alia*. Although the laboratory toxicity assessments may be deemed conservative, the potential for chronic effects in the field must be taken into account, even if the field data were "inconclusive."

2.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

Given the ability of frogs to range easily over one or two ponds, the estimates of exposure from sediments and food organisms is appropriate. However, the lack of toxicity information on salamanders (generally thought to be more sensitive than frogs) should lead to precaution in estimating HQs too closely in these studies. The variances in effect concentrations may be quite large.

2.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes!

2.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Yes. However, the use of log-linear models (chi-square) for tests of association are particularly sensitive to the individual cell sizes (Sokal & Rohlf, 1989; Quinn & Keough, 2002). An analogy can be made by thinking about an experiment in which the results are described as follows: "1/3 of the exposed frogs died, 1/3 of the exposed frogs lived, and the third frog got away!" The strength of the association between PCB soil/sediment concentration and the resulting reproductive or survivorship results depends not only on the number of comparisons (four ranges of PCB concentrations in this ERA), but on the number of occurrences in each "block" or "cell," as well.

2.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

In this reviewer's opinion, the determination of "high risk" to amphibians is supported by the toxicity evidence –particularly- and by the field evidence (lack of egg masses and/or adult females). In this instance, a precautionary approach would be dictated as well, for the frog species are probably less sensitive to PCB exposure than are the urodele salamanders.

The risk characterization would be enhanced (in my opinion) if the data in Figure 4.4-11 were to be put into an "SSD" format, as described above for benthic macroinvertebrates. The resulting statement of "90% [my estimate; not a calculation!] of amphibian effects endpoints responded to PCB concentrations of 10 mg/kg or less"

2.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

The uncertainties were addressed in the ERA. What was not as well addressed included data on urodeles. The uncertainty here is high – but unknown. Improvements would call for data specifically on salamanders or closely related species. This would probably best be accomplished under laboratory or "mesocosm-type" controlled exposures.

2.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

For the ranid species, the WOE was appropriate. For amphibians in the floodplain, probably not. However, this could only be made more robust by collecting more information.

2.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Yes. As the basis for most of the risk characterization is based on toxicity studies and malformations, the use of predictive approaches downstream are appropriate. The population modeling approaches (meta-population matrixes) are sensitive, when the matrix values closely approximate the seasonal means and variances in survivorship and reproduction. Again, basing the risk characterization on ranid frog species would appear to call for a conservative basis, as urodeles are strongly suspected of being more sensitive to PCB exposure.

2.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

Yes, in part. As a matter of professional judgement, the potential for sediment toxicity to larval frogs should be "moderate to high," given the high weighting, the demonstrated evidence of harm (albeit from laboratory toxicity studies) and the fact that the frogs are surrogates for salamanders, strongly suspected of being more sensitive to PCB exposure.

3. Survival, Growth And Reproduction Of Fish.

3.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

Yes. Electroshocking and nest surveys are readily accepted as methods. Further, the studies took pains to ensure that "sampling-effort" was standardized among the different field collections. The literature data, summarized in Fig. 5.3-1, is an excellent example of intensive literature-based analysis. With the results explained in Figs. F.4-7 and F.4-8, there is a clear indication of the TEQ threshold of 45 – 50 ng/kg and the 30 – 45 mg/kg threshold for tPCB.

3.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

Yes.

3.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

Yes.

3.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes.

3.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Yes!

3.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

Yes, in part. The risk characterization concluded "A high probability of adverse impacts to fish from tPCBs and/or TEQ…. [but mortality of adults is unlikely]. The magnitude of the risk is high, quite independent of adult mortality. Through the developmental toxicity testing, the literature data, and the field-measured limited recruitment into the bass population, the concluding risk characterization of "low" or "undetermined"magnitude is not, in my opinion, justified. The data are sufficient to determine at least moderate, if not high, magnitude. The fact that fish populations have been measured to be "self-sustaining" could be the result of the limited number of sampling events (= four) over the last 4 to 5 years. It may take longer for the full reproductive effects to show up. Further, the evidence of abnormalities may ultimately be a consideration for the overall population "health."

3.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

They were identified and addressed (particularly the sub-lethal effects in larval and YOY fish), but not appropriately weighted. In my opinion, the threshold effects levels should be established lower than they are presently. The effects on larvae and eggs (@ circa 11 mg/kg), determined from literature, should be deemed of sufficient strength to lower the 32 mg/kg threshold for warmwater fish. I would hope there is value in not broaching the "catastrophic" effect level before "high" magnitude is assigned.

3.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

No. The literature data and the toxicity exposures should be given more weight. From Figures 4-7 and 4-8, I would conclude that a lower threshold of circa 25-30 ng/kg TEQ and circa 30 mg/kg for tPCBs (Figures 4-1 through 4-5. Although it cannot be put "into evidence" for this ecological RA, it must be acknowledged that the conservative thresholds would also be protective of other higher predators in the Housatonic River valley, for example, *Homo sapiens.*

3.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Yes. The sediment characterizations and concentrations were such to show low risk.

3.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

Not completely, for the reasons outlined above.

4. Survival, Growth And Reproduction Of Insectivorous Birds.

4.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

By-and-large, yes. The three-year swallow study incorporated a more appropriate duration to study reproduction in this species. Given the high tissue PCB burdens, the weakly negative correlation between PCB body burden and hatching success (and the lack of clutch size relationship with PCB concentration), the risk conclusions should be given higher weight than that from the robin study. The one-year robin study was not of sufficient duration to quantify any effects on reproduction stemming from PCB exposure and it artificially adds uncertainty to risk estimates. Also, I continue to think that the ERA should give more weight to the swallows, as they are more directly linked to emerging insects from aquatic sediments. Although robins may be an appropriate surrogate species for bird species with a terrestrial food chain, the demonstrated PCB concentrations in the sediments and floodplain, and limited residues in upland areas, the focus should be on the swallows and other species relying on aquatic insects. In fact, the two species that could be selected to complement each other in terms of insectivory would be swallows and bats.

4.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

The three-year nest box field study for the swallows were highly appropriate. The use of next boxes (measuring the food intake, contaminant concentrations in diet, and body or tissue burdens) has been shown highly effective in other studies (Kendall, parathion study; Cobb-Hooper metal studies; McMurry, DDT studies). The three-year duration of the study was also approaching "sufficiency," as swallows live six to 8 years.

The one-year robin study appears to have several problems associated with it, not the least of which include the low statistical power stemming from next predation and the one-year duration. The variances provided with this study "cloud the issue" and do not appear to provide much quantitative information useful to the overall ERA for insectivorous birds.

4.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

Yes, for the swallow study. No, for the robin study. Overall, the use of micro-exposure models is laudable! In my opinion, this type of modeling makes the best use of a highly limited data set for contaminant uptake into insectivorous birds. Having said that, it should be recognized that each component of the model (e.g., food intake rate, PCB concentration in prey, maternal transfer) has it's own variance. In my opinion, carrying the variances through the modeling effort should be used to help justify conservatism in risk estimates and not to provide justification for lessening a risk characterization. The field study for the swallows was of minimally sufficient duration to support estimates of exposure obtained from the literature.

4.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes. Focusing on reproductive consequences and early life-stage consequences of PCB exposure is highly appropriate. Unfortunately, the study duration to quantify the effects needs to match the life history of the species under consideration. As has been quantified with exposure to other poly-chlorinated hydrocarbons, the "ultimate" effect may be observed in the offspring of nestlings exposed to high PCB levels.

4.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

The use of microexposure and daily intake modeling and the risk estimates from upper and lower risk bounds (and Monte Carlo simulations) are excellent approaches.

4.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

For the swallows, the characterization of risk, using a WOE approach, appears to be low. That is, the findings included high body burdens, weak – but statistically significant – negative correlation with hatching success, and measured uptake from food items. The only evidence of "no effect" comes from the lack of success on clutch size, a life history characteristic closely controlled by genetic factors. Given this evidence, and the limited value of the robin study, a ranking of "low risk" for tree swallows is questionable, in my opinion.

4.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

The uncertainties were very well taken into account. However, for these species, I would recommend that uncertainty add to a higher risk estimate, as the species are relatively long-lived and were only sampled over a limited period of their lives.

4.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

In my opinion, the WOE was not appropriate for these species. I might suggest using the variances associated with feeding, uptake, and maternal transfer and use the variance estimates to indicate the degree of uncertainty (it is acknowledged to be high in this ERA). As such, the WOE approach may not be as useful to estimate long-term results, stemming from measures taken on an "acute" (one year for the robins) or "limited" (three years for the swallows) basis. The only way to improve the analysis is to re-do the robin study, for a minimum of two or three years, perhaps including banding adults and their offspring. Also, if a limited number of swallows were to be banded and studied over three-to-four years, it would help measure effects on the reproductive success of the F1 generation.

4.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Given the limitations noted above, yes, they were objectively applied.

4.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

Please refer to my response in 4.h. No, I do not think the conclusions of low risk to tree swallows are supported by the results. I would have concluded "moderate-to-high" risk, after reviewing the data. I have not put much weight (either way) on the robin study. In my opinion, that study should form the basis for a second study on robins, conducted over a minimum of two or three years.

5. Survival, Growth And Reproduction Of Piscivorous/Carnivorous Birds.

5.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

Given the limited amount of data for piscivorous birds, the use of kingfishers and osprey was valuable and appropriate. The kingfisher field study, albeit conducted with the best of intentions, has very low sample size and hence a low "power of the test." As with the insectivorous birds (in Section 4, above), the life span of kingfishers and osprey are very long compared to the duration of the field study. Hence, effects would not be expected to have been measured, unless they were of an extreme, acute, nature. However, the studies were certainly conducted based on accepted scientific practices.

The EPA modeling studies, using literature data and threshold values, are also based on accepted scientific practices.

5.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

See my response under 5.a. Conducting a study of this magnitude takes a lot of effort, time and personnel. However, for a relatively long-live species, to determine reproductive or growth effects with a one-season study produced results of limited value. Monitoring the burrows for kingfishers is laudable! However, very little, if anything, can be concluded from the results of six non-depredated clutches. This field study should be given little weight in the overall ERA.

5.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

Yes. I think the use of fish –rather than the limited data available for crayfish – was appropriate. In other situations where crayfish are more common, their use as prey items would be more important in estimating dietary intake. Given the nature of the Housatonic, and the fish community, the “substitution” of fish for crayfish is acceptable.

5.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

The effects metrics identified were appropriate. Whether or not they were adequately assessed is questionable. The field study has limited use in this circumstance.

5.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Yes.

5.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

Yes, the characterization of risk was appropriate for the kingfisher, based on the modeled exposure and effects. The variance and “clouding” of results provided by the admirably attempted field study does not provide information useful in determining effects on the kingfisher population. The field study would either have to have been conducted over a longer time period, with marked individuals, or have been conducted during the one season with more than six clutches. I recognize the difficulty in conducting such field experiments and also recognize the ease of criticizing them; I am not in any way minimizing the effort it took to gather the data. Nor am I denigrating the work conducted: it was conducted with the best of intentions. However, because of the limited number of burrows, and then some of those burrows depredated, not much can or should be made of the results.

All this supports my conclusion that the WOE risk categorization of “high” evidence of harm should hold for both species and overall for Piscivorous birds.

5.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

Yes, the uncertainties were acknowledged. The only improvement I can suggest is to include other –cageable- piscivorous birds (such as mergansers) and, if another field study were to be performed, to recommend that the study duration be as long as possible, well over two years and using banded birds. The use of a surrogate species, like a fish-eating duck, to directly study PCB uptake would lessen the need for prey PCB burden estimates. Also, should the caged ducks be studied over one reproductive season, the maternal burden provided to eggs or nestlings could be accounted for.

5.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

Please refer to my response under 5.f., above. I do not think the WOE analysis is appropriate for these species, as the study duration was not long enough, if the field study is to be given weight in the ERA conclusions. The modeling study provides a more conservative estimate, yes, but this (and the conclusions stemming from accepting the higher degree of risk) should be weighed against the loss of the avian resource. In Table 8-5.3, in Volume 2, Section 8, the modeling results indicate risk is “high” for the kingfisher and osprey. The weighting of the results (Tables H 4-6 and 4-7) indicate “moderate weighting.” The field study indicates “low” risk. However, the field study, unfortunately, is simply not useful in determining long-term effects.

5.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

As there are no data for these risk estimates, the modeling results should hold (as “high” evidence of harm, with moderate weighting). As uncertainty in the data provide for a conclusion of “low” risk, I disagree with this conclusion. Uncertainty, with these two species and given the limited information on piscivorous birds, should lead to more conservatism.

5.j. In the Panel members’ opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

No, for the reasons detailed in Sections 5.h, I, and j, above.

6. Survival, Growth And Reproduction Of Omnivorous/Carnivorous Mammals.

6.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

The strongest studies in this component were the modeling studies, using literature (rat) data and the GE demographic field study (with re-analysis). However, all studies indicate an extremely large degree of uncertainty, which precludes any strong statement of harm or risk. However, that said, the studies that were conducted did follow accepted scientific procedures. The semi-quantitative study of shrews collected in the amphibian study are weak – and should lead to a conclusion that a study focused on omnivorous field mice, oppossums, or raccoons (using radio collars) would have been beneficial. Yet, that is hindsight…..

6.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

With the re-analysis of original results, the GE short-tailed shrew study was very strong and indicates a high level of potential harm (at least in Location 13) to insectivorous small mammals. Given the well-published consequences of PCB exposure on mortality and reproductive effects in mammals, the results from this study that relate soil PCB concentration to mortality (albeit with broad confidence limits) should carry high weight.

6.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

Yes, the spatially-weighted estimates for shrews was particularly appropriate, as shrews have such limited daily ranges. The estimates for the red fox are much less quantitative and have huge uncertainties associated. However, the uncertainties were acknowledged in the ERA. In my opinion, the re-analysis of the GE shrew demographic study provides some excellent trend data towards demonstrating consequences of exposure to soil-bound PCBs.

6.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes. The emphasis on reproductive effects is good. Given the quantified mortality in the GE shrew study, and with the literature on rats indicating substantial reproductive effects at low-level, chronic exposure to PCBs, this metric is valuable.

6.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Including the re-analysis of the GE shrew survey, yes.

6.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

Results from modeling exposure and effects (largely using rat data) suggest that PCBs in soils pose a high risk to short-tailed shrews inhabiting Locations 13 and 14. In my opinion, much less can be said of the “intermediate risk” to red foxes exposed to tPCBs and TEQ in the PSA. The large uncertainty concerning the modeling line of evidence for tPCBs and TEQ forces the conclusion that risks of these COCs cannot be really estimated for the PSA. The GE shrew survey (with re-analysis) indicates that the small, insectivorous and/or omnivorous mammals are at risk. However, there may be extensive immigration into the PSA. This is simply unknown at present.

6.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

There is so much uncertainty in this component (insectivorous and carnivorous small mammals) that conservatism must be called for, unless further emphasis is placed on quantifying the distribution and behavior of red foxes and other small mammals (raccoons, opossums, and the abundant number of mouse species). In this analysis, as with the Piscivorous mammals, above, a detailed study of red foxes, with radio collars, would be invaluable.

6.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

This is a particular instance, in this reviewer’s opinion, where the WOE approach does not work. When all avenues of evidence are highly “suspect” because of limited quantification, limited sampling success, and high uncertainty, the WOE approach should have an option to “go for more data.” As it is, given the large number of mouse and shrew species, given the negative regression of shrew mortality and soil PCB concentration, given the well-published results indicating reproductive effects stemming from chronic PCB exposure in small mammals, I would recommend not putting much weight on the WOE approach and employ the results of the Monte Carlo and probability bounds modeling and the limited – but insightful – results of the GE field study.

To lessen risk estimates because of so much uncertainty is neither wise nor prudent. This situation calls for either more data or more conservatism.

6.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Yes, but the risk estimates are so broad as to not be very useful. The limited data (literature and field study re-analysis) would indicate a greater risk than indicated by the WOE approach. Given the large number (40+) of small mammal species on the PSA, conservatism at Locations 13 and 14 would seem to be called for.

6.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

No. See my responses to 6g., 6h., and 6i for justification of my opinion.

7. Survival, Growth And Reproduction Of Piscivorous Mammals.

7.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

Yes. The mammals selected for study, mink and otter, are very difficult to monitor in the wild. Hence, the reliance on modeling parameters and the mink feeding study (MSU) were highly appropriate. Although useful, the wide variances associated with locating, tracking, and identifying mink or otter during winter has less use – but was appropriately conducted.

7.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

Yes. The survey for mink and otter tracks (including scent posts, visible sightings, and foot track identifications) were appropriate. However, as with most initial surveys, the surveys provide good background for how to conduct further studies. Such studies, were they to be conducted, might include collaring selected adults and following them, incorporating GIS approaches to determine home range, migratory range, and den location(s). Similar studies (of radio-collared animals) have been conducted on mink and marten in the Wyoming and Colorado Rockies, with considerable success.

7.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

The exposure period appears to account for seasonal variability in dietary composition for mink and perhaps less so for river otter. There is great value in using an exposure period that encompasses the reproductive cycles of these species. Although experimental, and controlled by the laboratory diet, results of dietary exposure were also well assessed in the mink feeding study. From Section 9, Volume 2 and Appendix I, it appears that the estimates of dietary composition (percentage of fish, invertebrates, etc) and daily intake were appropriate. One particularly good approach was to incorporate estimates of 10% and 100% foraging times in the PSA. These estimates provide a good "bound" for modeling exposures.

There is much more uncertainty in the estimates for the otter. However, without including a radio-collared study of otter migration, home range, and "local" behavior patterns, there will remain a large uncertainty (as expressed in this ERA) on otter exposure.

7.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes. The feeding study is particularly strong and results stemming from that study should (in my opinion) carry a lot of weight for the ERA results for piscivorous mammals. The studies cited in Appendix I indicate that mink are very sensitive to dietary PCB levels, confirming results of the feeding study in this ERA. There were, of course, no effects metrics quantified by the surveys; hence, these field studies were of much more limited use in determining effects on mink or otter. The fact that no mink were found (EPA study) in habitats appropriate for mink is somewhat disconcerting, and somewhat modified by the GE field survey. However, both studies are field surveys for species known to be extremely difficult to monitor in the wild, without radio collars attached. Hence, one would expect (at least this panelist would expect) few quantitative results on effects metrics to derive from field surveys.

The USEPA may gain some valuable insights into the NOEL for "aquatic mammals" in the recently-published (2001) NAS/NRC publication, "A Risk-Management Strategy for PCB-Contaminated Sediments." In it (pg.405), an NOEC for tissue residue total PCB was estimated to be 10   g/g, lipid normalized. The number was cited from the work of Kannan, et alia, (2000).

7.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Yes. The feeding study was very well conducted and used appropriate controls. The results, well-quantified, are strong and telling.

7.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

Somewhat. The field surveys (either EPA or GE) have large variances in sampling and quantifying mink or otter use. This is completely understandable for such types of species as mink and otter. They are difficult to study in this manner. The primary results (feeding study and literature survey) should have very high weighting value – much higher than the field surveys. Given that, the risk rating should remain "high," not "intermediate to high."

7.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

Yes, they were adequately addressed and acknowledged to be high. However, given the extensive literature data on the sensitivity of mink to PCB exposure, the uncertainty should be treated with conservatism in assessing risk to mink or otter. In my opinion, the only improvement would be to incorporate a radio-collar study of five to 10 mink and follow them over one or two years. The cost of such a study would be high – but relative to the value of the resource, the cost would be comparatively small.

7.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

The WOE analysis "overweighted" the field study, in my opinion. There is sufficient literature and, when included with the mink feeding study, the weighting on the modeling efforts should be increased, leading to a decision of "high risk" for Piscivorous mammals feeding within the PSA.

7.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Please refer to my responses in 7.g. and 7.h. In my opinion, the risk estimates as provided in this ERA (= "moderate to high") were derived by providing too much weight to the field studies.

7.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

No, not strictly. I think the evidence points towards a greater risk for piscivorous mammals than is presented in this ERA.

8. Survival, Growth And Reproduction Of Threatened And Endangered Species.

8.a. Were the EPA studies and analyses performed (e.g., field studies, site-specific toxicity studies, comparison of exposure and effects) appropriate under the evaluation criteria, and based on accepted scientific practices?

Yes. There was an emphasis on modeling, as no bald eagles or bitterns were observed during the surveys, despite having observations of their presence seasonally.

8.b. Were the GE studies and analyses performed outside of the framework of the ERA and EPA review (e.g., field studies) appropriate under the evaluation criteria, based on accepted scientific practices, and incorporated appropriately in the ERA?

N.A. The field study conducted was a survey to look for the presence of eagles, bitterns and s.f. myotis (among other T&E species).

8.c. Were the estimates of exposure appropriate under the evaluation criteria, and was the refinement of analyses for the contaminants of concern (COCs) for each assessment appropriate?

The use of literature data and the concentrations in potential prey items was appropriate.

8.d. Were the effects metrics that were identified and used appropriate under the evaluation criteria?

Yes.

8.e. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

Yes. This was a good use of probabilistic modeling.

8.f. Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

Yes.

8.g. Were the significant uncertainties in the analysis of the assessment endpoints identified and adequately addressed? If not, summarize what improvements could be made.

Yes. The large uncertainties arise from the lack of specific PSA data (presence, feeding, territory size and use) on any of the T&E species.

8.h. Was the weight of evidence analysis appropriate under the evaluation criteria? If not, how could it be improved?

Yes. One way to improve the WOE analysis would be to band several eagles and/or other T&E species and follow them for several years. Another way to improve the WOE would be to use surrogate species (mergansers) to study PCB and TEQ uptake from the fish in the PSA. However, please note my comment in 8.j.

8.i. Were the risk estimates objectively and appropriately derived for reaches of the river where site-specific studies were not conducted?

Somewhat. If the osprey is considered at high risk, but uncertain because only one line of evidence was employed, why not the same conclusion for the bald eagle and bittern? Ospreys and bald eagles are competitors in habitats in which they coexist. Hence, a logically-applied WOE would conclude very similar risk levels for both species.

8.j. In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local population of ecological receptors?

Yes. I note that in this instance, with large charismatic T&E species, that the WOE approach fully uses and relies on modeling. If the same approach were to be used for piscivorous and omnivorous mammals, the reliance on limited field data would be lessened and greater weight put on the modeling. If the same "weighting" on high-variance, small sample field testing were to be applied here, it would moderate the "moderately high" risk determination by lessening it to "moderate." Could it be that shrews and foxes are not held in as high esteem as bald eagles?

4. ARE THE SUMMARY DISCUSSIONS AND CONCLUSIONS IN THE ERA SUPPORTED BY THE INFORMATION PROVIDED IN THE REPORT, AND DID THE CONCLUSIONS DESCRIBE THE RISKS IN AN OBJECTIVE, REASONABLE AND APPROPRIATE MANNER?

Mostly. There are inconsistencies in approaches for the piscivorous birds and the T&E representative, the bald eagle. As an idea and in my opinion, the bald eagle (or mink) results could be used as a surrogate measure for the ospreys, kingfishers, and other predominately fish-eating vertebrate fauna. The ERA acknowledges that there are substantial differences in species sensitivities to COCs (particularly the PCBs). However, given the value of using surrogate species, species chosen to represent other species present in the PSA, care should be taken to be conservative, as it is most certain that several un-measured species are more sensitive than are the ones selected to represent the biotic community (e.g., data presented in Table 12.4-1).

In the Conclusion Section 12.2.1.3 ("Fish"), the statement concludes, "The magnitude of impact is not predicted to be catastrophic in any reach;" I would propose that the manner in which PCBs exert their toxicity is not in a "catastrophic" manner, but in chronic effects on reproductive behavior. Further such reproductive effects will ultimately influence the survivorship of predators feeding on fish in the PSA – hence, conservatism should be called for. Subtle responses require more samples, longer study durations and truly focused studies on the critical questions ("would a full life-cycle test with trout or bass indicate reproductive effects in the F1 or F2 generations?").

In almost every situation (insectivorous or carnivorous birds and mammals, for example), the field studies provided uncertainty in this ERA. This is not to say that those studies were poorly planned or conducted. They were good efforts to analyze population responses. However, as it is acknowledged that PCBs produce their affects over long-term exposure, the studies must have a duration and "intensity" to be able to quantify the subtle responses (a good example of a sufficiently-long study was the three-year swallow study). As evidenced by the complete reliance on modeling results for T&E species, the modeling results almost invariably result in estimates of greater risk. In my opinion, adding "clouding" and variance to these modeling estimates does not further the protectiveness of the risk assessment, particularly when the cost of conducting the focused studies is compared to the value of the species to be protected.

My comments above are particularly germane under the "Protection Goals" Section of the ERA (Section 12.4.2.1), in which broader aspects of the biotic community are to be protected as an ERA goal. Hence, the value of the mink, frog, osprey and eagle endpoints are not limited to protection of each species population, alone, but also as surrogates for the other species in the PSA. This section also calls for conservatism, as I would state we are not sure where "overly-conservative" estimates come in – but it is better to err closer to "over-conservative," than to err on the "upper end" of concentration limits.

As the ERA states (Pg. 12-70; line 31ff: "the central question….. is whether the exposed local populations are thriving in the contaminated habitat, not whether the larger regional population is surviving in spite of it." Because of the subtle, and decidedly un-catastrophic nature of the effects of PCBs, it must be kept in mind that studies of such effects must include a combination of extended duration, focused experimentation (feeding, collaring, banding, or full life-cycle testing) to be able to have any success in determining the boundary effects levels, including solid estimates of NOECs and LOECs.

5. **TO THE BEST OF THE PANEL'S KNOWLEDGE, IS THERE OTHER PERTINENT INFORMATION AVAILABLE THAT WAS NOT CONSIDERED IN THE ERA? IF SO, IDENTIFY THE STUDIES OR DATA THAT COULD HAVE BEEN CONSIDERED, THE RELEVANCE OF SUCH STUDIES OR DATA, AND HOW THEY COULD HAVE BEEN USED IN THE ERA.**

The only consideration I would strongly suggest is to add two things to this ERA: 1) A section not unlike the Executive Summary, but one that includes an overview from the ecological perspective. Although there are to be individual endpoints, some consideration should be given to risk (as for the benthic invertebrates) carried "up" to the fish. This consideration should also take into account the "health" of the population. The Framework for Ecological Risk Assessment focuses on populations. Hence, a discussion of top predators must include "lower tier" assessments of their food webs. It follows that, if the prey organisms are at risk, ultimately, so must the predators be. 2) For the benthic invertebrates, there are several good multivariate approaches that may link the multiple variables to benthic invertebrate abundance and distribution. Such citations include:

REFERENCES CITED:

Green, R. H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley & Sons, NY. 257p.

May, R.M. 1975. Patterns of Species Abundance and Diversity. In M.L. Cody and J.M. Diamond (eds.). Ecology and Evolution of Communities. Pgs. 81-120; Belknap Press, Cambridge, MA.

Moriarty, F. 1999. Ecotoxicology, 3rd Edition. Academic Press, NY.

National Academy of Science. 2001. A Risk-Management Strategy for PCB-Contaminated Sediments. (page 405) National Academy Press, Washington, DC.

Posthuma, L., Suter, G.W., II, and Traas, T.P. 2002. Species Sensitivity Distributions in Ecotoxicology. Lewis Publishers, Boca Raton, FL. (Chapt. 9, pgs. 155 – 193).

Quinn, G.P. and Keough, M.J. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge, UK. (Chapter 14, pgs. 380 – 400).

Tabachnick, B.G. and Fidell, L.S. 1996. Using Multivariate Statistics, 3rd Ed. Harper Collins College Publishers, NY. (Chapter 6, pgs. 195 – 238)

Van den Brink, P. J., and Ter Braak, C. J. F, Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. *Environ. Tox. and Chem.*, 18, 138, 1999.

Van Wijngaarden, R. P. A., Van den Brink, P. J., Oude Voshaar, J. H. and Leeuwangh, P., Ordination techniques for analyzing response of biological communities to toxic stress in experimental ecosystems, *Ecotoxicology*, 4, 61, 1995.