Housatonic River
Rest of River Ecological Risk Assessment
Review Panel Comment Submission Form

Name of Panel Member Completing Form: James T. Oris, Ph.D.
Date: 16 February, 2004

A. Introduction:

In considering these questions, the Panel members shall evaluate the following (hereinafter the "evaluation criteria"): the objectivity, consistency, and reasonableness of both the procedures and inputs used by EPA in the application of existing EPA guidelines, guidance, and policy; and those used by EPA in the absence of Agency guidelines, guidance, or policy (see Attachment A for the list of relevant EPA guidelines, guidance, and policy documents).

If significant errors are observed in the application of the appropriate methodologies, the Panel members shall provide specific comments, describing the error(s) and suggested improvements. The suggested improvements must be specific, clear, and consistent with existing EPA methodologies and guidelines.

Format of Comments: Document/Chapter/Line number (or Figure or Table number); Panelist Comments; Proposed Change.

B. General Comments about the documents:

The ecological characterization and the ecological risk assessment presented for the Housatonic River ecosystem is a very impressive set of documents. At the surface, they are extremely comprehensive and thoughtfully compiled. However, there are numerous aspects of the organization of the documents that should be considered prior to finalization of the ERA. Two areas of revision would be considered major revisions to undertake.

Comment B1. The first area is of document organization. Currently the document is extremely large and contains an enormous amount of information. It would serve the Agency, GE, and the public if a significant effort were exerted to organize the document better. The document is organized with a general executive summary, specific chapters with more details, and detailed appendices for each chapter, and then detailed appendices that cover methods and analysis that overlap among chapters/appendices. Based on my reading, the specific chapters and the appendices have a tremendous amount of overlap and redundancy. This was explained as being necessary because the document was meant to serve multiple purposes and audiences. I submit that most laypersons will only read a 1 - 10 page brief summary (e.g., as in the glossy documents meant for the public). Many highly interested persons with a variety of educational backgrounds will be able to work through and understand the executive summary. In my opinion, only technically competent people will read the specific chapters or be interested in the appendices. Therefore, I see no reason to have both chapters and their associated detailed appendices. I submit that the document would be best organized by having all of the detail within the specific chapters. Those persons reading the specific chapters will be, because of their professional training, accustomed to skimming sections for appropriate detail. They also (like myself) will be frustrated by repeated cross-referencing to different parts of the documents, and then having to search through redundant information to find the one or two paragraphs that have those details. A 'redundancy analysis' or 'compare documents' exercise between specific chapters and their appendices will show that a remarkable reduction in document length and a remarkable increase in organizational clarity could be achieved by putting the details in the chapters and eliminated most of the appendices.

This could be achieved with a minimal disturbance to the readability of the document. There are some appendices that should not be eliminated, however. These would include sections that are redundant among chapters (e.g., methodology on setting values for non-detects).

Comment B.2. The second area is of document transparency. I found it difficult to connect all the pieces of the ERA together into one coherent overview or conceptual framework. In addition, a significant effort needs to be exerted to eliminate confusing statements (e.g., see comment 1.1 below), differential application of methods (e.g., risk characterization of fish and invertebrates versus birds and wildlife), and differential use of terminology (e.g., risk terminology). The organization of the document (as outlined above) also hinders the transparency of the document (it is hard to find information you are seeking because it is in so many different places). Finally, the least transparent portion of the document is the description and application of the weight of evidence approach. How final weightings were assigned, how each line of evidence was used, why sometimes lines were eliminated, and how final rankings were decided based on all lines and weights combined is not clear in any of the sections. This is the case even after reading the weight of evidence papers in the literature. Perhaps this is just my own confusion, but I think it would be useful to provide a detailed, step-by-step example of how the method was utilized for one of the receptors and then ensure that the method is applied consistently throughout the document. The weight of evidence approach, with the current level of explanation, is open (in my opinion) to criticism of subjectivity in the conduct of the risk assessment. I do not believe that there was bias in judgements by EPA, but this is an area that can be considered a weak point in the document.

An example of the WOE unclarity can be found using an analysis of the assignments of weights in Table D.4-1 for the invertebrate assessment. Below is a reproduction of that table, objectively assigning point values to each weighting classification, using three different averaging methods to calculate weighting classification scores, and then a comparison to the EPA assigned overall weightings in the original table. , While there is detailed explanation for why each item in the table received a particular weighting, the lack of transparency in the EPA approach is that there is insufficient explanation of how the final, overall endpoint weighting value is assigned. In my analysis below, the first averaging method assumes each category of weighting values has equal weight, calculates an average within each category and then an overall average of the three categories. The second averaging method assumes that each of the 10 subcategories has equal weight and the average is calculated over all of these values (i.e., sum of raw weightings over a column divided by 10). The third method calculates average weighting values within a category, re-ranks the weight, and then averages the value of the new ranking. Either method 1 or method 3 are probably the most appropriate because they provide even estimates of weights across the three categories since there are different numbers of subcategories within the three (i.e., Relationship between Assessment and Measurement endpoints (3 subcategories), Data Quality (1 subcategory), Study Design (6 subcategories). Regardless of how these are calculated, there is a fair amount of discordance for several of the endpoint group weightings between my analysis and the EPA-assigned values. This is especially true for endpoint groups C1, C2, C3, and T3 (nearly 50% of the endpoint groupings). I am not sure how this will affect the outcome of the risk characterization for invertebrates, and I did not take the time to evaluate each of the weighting tables. It is evident, though, that there needs to be more objectivity, better explanation, and more clarity to this analysis than is currently present in the document.

Analysis of Table D.4-1, Weighting of Measurement Endpoints.

Rankings within single cells: 1=Low, 2=Low/Mod, 3=Mod, 4=Mod/High, 5=High
Rankings of averages: 1-1.66=Low, 1.67-2.66=Low/Mod, 2.67-3.66=Mod, 3.67-4.66=Mod/High, 4.67-5.00=High
Abbreviations: L=Low, LM=Low/Mod, M=Mod, MH=Mod/High, H=High

Category:	Endpoint	Groups:							
	C1	C2	C3	T1	T2	T3	B1	B2	B3
Meas vs. Assess Endpoint:									
1	1.00	1.00	3.00	4.00	4.00	3.00	2.00	2.00	2.00
2	1.00	2.00	3.00	4.00	4.00	4.00	2.00	2.00	2.00
3	1.00	3.00	3.00	5.00	5.00	4.00	2.00	2.00	2.00
Avg of Cat.	1.00	2.00	3.00	4.33	4.33	3.67	2.00	2.00	2.00
Rank of Cat.	L	LM	M	MH	MH	MH	LM	LM	LM
Score of Cat.	1.00	2.00	3.00	4.00	4.00	4.00	2.00	2.00	2.00
Data Quality:									
4	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Avg of Cat.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Rank of Cat.	H	H	H	H	H	H	H	H	H
Score of Cat.	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Study Design:									
5	2.00	2.00	2.00	4.00	4.00	4.00	5.00	5.00	5.00
6	2.00	2.00	2.00	5.00	4.00	5.00	2.00	2.00	2.00
7	4.00	5.00	3.00	3.00	4.00	1.00	4.00	4.00	4.00
8	5.00	5.00	4.00	3.00	3.00	3.00	3.00	3.00	3.00
9	4.00	4.00	4.00	5.00	5.00	4.00	5.00	4.00	3.00
10	3.00	5.00	5.00	5.00	4.00	4.00	3.00	5.00	5.00
Avg of Cat.	3.33	3.83	3.33	4.17	4.00	3.50	3.67	3.83	3.67
Rank of Cat.	M	MH	M	MH	MH	M	MH	MH	MH
Score of Cat.	3.00	4.00	3.00	4.00	4.00	3.00	4.00	4.00	4.00
Avg of (Avg of Cat.)	3.11	3.61	3.78	4.50	4.44	4.06	3.56	3.61	3.56
	M	M	MH	MH	MH	MH	M	M	M
Overall Avg-no Cats avg'd	2.80	3.40	3.40	4.30	4.20	3.70	3.30	3.40	3.30
	M	M	M	MH	MH	MH	M	M	M
Avg of Cat. Score	3.00	3.67	3.67	4.33	4.33	4.00	3.67	3.67	3.67
	M	MH	MH	MH	MH	MH	MH	MH	MH
EPA Rating in D.4-1	LM	LM	M	MH	MH	M	M	M	M

C. Charge Questions and Comments:

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?

Some of the best work done in the ERA was the ecological characterization portions. The document describing the ecosystem of the Housatonic River watershed is very thorough and clear with few exceptions.

Comment 1.1. Appendix A.1, section 1.1: Purpose of document states 8 work plans were addressed. Of these workplans listed, there is no mention of fish, macroinvertebrates (other than dragonflies), or soil infauna. Yet, there are entire chapters devoted to each of these groups in the document. Further explanation at this point in the document would clarify the role of the groups not included in the listed workplans.

Comment 1.2. Appendix A.1, Attachment C, Species matrix. What does "SC" mean in species status? Also, there is no analysis of this data (at least for fish) for community quality. For example, for the fish community, Attachment C indicates that 49% of the suitable habitat in PSA was not occupied with expected fish species. I believe a fish community analysis would be appropriate beyond just the description of what is there.

Comment 1.3. It is not clear whether all the studies in the ecological characterization were utilized in the ERA. Specifically, the dragonfly data (which is quite robust) is apparently not used in the ERA.

Comment 1.4. An inventory of data collected and used for both abiotic and biotic characterization is necessary to clarify what is present in the document. This will help in the presentation of the document and help clarify the decision processes used in the problem formulation phase.

Comment 1.5. The Massachusetts portions of the river have been extensively characterized, but the Connecticut reaches have only been treated superficially. This needs to be justified in the document, or further characterization needs to be done.

2. Was the screening of contaminants of potential concern (COPCs), selection of assessment and measurement endpoints, and the study designs for these endpoints appropriate under the evaluation criteria?

Comment 2.1. Screening of COPCs. After hearing and reading answers to questions posed to EPA by the review panel, I believe that the screening process was appropriate. However, the specific sections that describe the screening process and the decisions for what to include in the COC list need to be revised to clarify previous questions from the panel.

Comment 2.2. Selection of assessment and measurement endpoints.

2.2.1 As I have previously stated in my questions, I do not quite understand the wording of the assessment endpoints. Assessment endpoints are supposed to encompass the goals of the risk

assessors. Stating an assessment endpoint as "Survival, growth and reproduction of birds" doesn't' tell me what the desired outcome or goal is to be. The response to my question about this form of statement was that the goal was no impairment of the stated endpoints. A goal of "no impairment" is probably unrealistic given the multitude of potential stressors in a combined industrial, urban, agricultural, and rural watershed. "No impairment due to PCBs" could be a realistic endpoint. "No impairment relative to reference conditions" could be a realistic endpoint. Regardless of what the sidebar-stated goals are of these assessment endpoints, they are not enumerated in the ERA document. The statement "Survival, growth, and reproduction of fish" is a value-neutral statement and is not a well-defined assessment endpoint. This is more a statement of a measurement endpoint to me. In the documents presentation meeting, the EPA's presentation by Sue Svirsky used the following statement for the assessment endpoints (slide 10 of presentation): "Survival, growth and reproductive success of [list of receptors]". Note the word "success" is in this statement and doesn't appear in the ERA document. Additional explanation or revision of the assessment endpoint statements in the document is necessary.

2.2.2 It is not clear why some endpoints were more restrictive than others. For example, community structure is a stated endpoint for invertebrates, but not for fish. This is even though fish community structure was analyzed and discussed in the document. In addition, population structure modeling was an endpoint for amphibians, but not for fish. This is even though fish displayed an abnormal age structure and there was evidence of reduced reproduction, and there was apparently little effort devoted to identifying areas of recruitment. In addition, no attempt was shown of field-level assessment of fish health (community or population) with PCB concentration, even though tissue residue data were available and DELT data provided with my questions indicate that there were high levels of deformities, some of which correlated to deformities seen in laboratory studies in a dose-related manner with PCB concentration.

2.2.3 All assessment endpoints need to be re-evaluated in terms of the societal value of the receptors. The decision to use "population-only" level effects (i.e., a reproducing population) may not be appropriate for all receptors. EPA guidance allows for the incorporation of societal value when appropriate. In this risk assessment, this is especially evident with the sections dealing with fish. It is apparent that society places high value on the health of fish. I differ in opinion with the EPA and others on the panel as to what a population level effect is here. To me, a high incidence of disease in a group of fish is a population level effect. As evidenced by public comments -- every single public speaker at the Cranwell meeting said the same thing-- it is clear that the public considers a diseased population of fish (regardless of whether they are reproducing) to be impaired. It is also clear that, even though (as stated by the EPA) the State and Federal agencies that spoke at the meeting participated in the development of assessment endpoint, these same agencies and the public felt very strongly that the fish assessment endpoint was incorrectly defined.

Comment 2.3. Study designs appropriate? I do not agree with some of the restrictions that the stated assessment endpoints generated in the design of studies (cf., comment 2.2.2 and 2.2.3), but generally, within the stated assessment endpoints, studies within the PSA were appropriately designed. There are some studies that could have been done differently or some analyses that still may be possible given the designs (e.g., fish deformities v. PCB tissue residue), but overall they appear to be adequate.

3. For each of the 8 assessment endpoints evaluated in the ERA (listed in Attachment B, and for which a specific Section and Appendix was prepared), address the following questions (discuss and label responses as 3.(assessment endpoint number).(question letter) for consistency):

NOTE: Because of the large amount of redundancy between chapters and appendices, and because more detail is presented in appendices, my comments will reference pertinent sections in the appendices only. Appropriate linkages will need to be made between appendix and chapter.

(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?

(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?

(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?

(d) Were the effects metrics that were identified and used appropriate under the evaluation criteria?

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

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

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

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

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

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

3.1 Benthic Invertebrates

3.1(a).1. In general, studies and analyses were done as described and meet accepted scientific practice.

3.1(b).1. No GE studies are presented in the ERA for invertebrates. However, the EPA needs to address concerns raised by GE's reanalysis of data.

3.1(c).1. Estimates of exposure and the COC refinement was appropriate. However, there seems to be some discrepancy between TOC and PCB concentration relationships with no explanations. In addition I am confused about the relationships between chironomid abundance versus taxonomic diversity related to substrate (fine, coarse) and tPBC concentration. These relationships appear to be opposite of one another (slides 36 and 38 in invertebrate presentation from document presentation meeting). This difference indicates that general conclusions about invertebrate toxicity in fine versus coarse sediment related to tPCB cannot be made. Sensitive species (chironomids) seem to have higher rates of toxicity in fine sediments, but species diversity seems to be more sensitive to tPCBs in coarse sediments. This will require further discussion and deliberation.

3.1(d).1. Development of MATC values seems somewhat subjective. However, it was good that multiple endpoints (EC20 and EC50) were evaluated. Would it not be better to use cumulative frequency distributions of EC20's or EC50's and choose MATC based on a measure of central tendency from the distribution rather than the subjective description used in the document?

3.1(e).1. The statistical approaches used in the invertebrate section are not always clear and are not fully explained in the document. It is not possible to evaluate whether statistical techniques were properly applied in all cases since insufficient detail is present to make such an evaluation. Analysis of statistical power, when appropriate, is not provided and queries from panelists concerning power analyses were not adequately addressed. The approach described in Appendix C.2 for handling non-detects would appear to be appropriate, but it is not clear that the procedure was used in all cases. Sections in both fish and invertebrate assessments refer to Appendix C.2, but then unless I missed something, non-detects in the invertebrate assessment were set at 0.5 of the detection limit and were set to the value of the detection limit in the fish assessment. In both assessments then, it appears that fixed-replacement was chosen (with little evidence to support the decision), but two different replacements were used. There seems to be a disconnect between the two assessments and it is not clear how these values were chosen.

3.1(f).1. The general characterization of risk is supported by the available information and appears to be appropriate. However, there are several areas of needed improvement in the document and these are included below as separate comments.

3.1(f).2. Risk terminology used within the invertebrate section needs to be revised and brought into concordance with other assessments (e.g., wildlife sections). Based on current terminology, any HQ value less than 1 presents no potential risk. HQ values greater than one present varying levels of risk on what seems to be a scale of HQ=1-3 (low risk), HQ= 3-10 (moderate risk), and HQ>10 (high risk). However, this scale is neither formally defined nor justified in the document, and the scale seems to be arbitrary. Based on the methods of developing benchmarks (non-conservative) and the use of high-levels of effect, all HQ values greater than 1 should indicate significant potential for risk. Levels of HQs that approach 1 may also indicate risk potential. For example, there is little possibility that there is a difference in risk potential between an HQ=0.9 and HQ=1.1, yet under the current scenario, the HQ=0.9 is considered "no risk".

3.1(f).3. Application of HQ values in the risk characterization is wholly deterministic, even though a probabilistic treatment could be used (such as in wildlife assessments). It is strongly suggested that the risk terminology and the risk characterization be conducted using the

approach used in the wildlife sections. This approach, defined using exceedance probabilities will be more appropriate.

3.1(f).4. A comparison of section D.4.2.1.3 (other COCs) and the data (Figs. D.4-1 and D.4.-2) indicates a high potential for risk due to PCBs, but downplays the potential importance of risk due to other COCs. The figures and the description of the HQs in the figures states that many of these COCs had values greater than 1 and less than 10, yet these potential risks are described with judgmental language that downplays the potential risk. Barium, chromium, copper, lead, mercury, silver, and tPAH all had median HQs > 1 at more than one site. Mercury had an upper range HQ of close to 100 at one site. Only median HQs were considered. The potential risks due to other COCs were downplayed based on the relative values of HQs between PCBs and other COCs. However, I submit that tPAHs and several of the metals listed above (esp. Hg) can serve as additional significant sources of risk. The HQ approach used here cannot separate out relative effects of PCBs and other COCs. Once an HQ is greater than 1, there is a significant potential for risk. To say that other COCs were not important because their HQs were "barely over 1" and "most often less than 10", is not appropriate. In addition, this approach cannot assess additive or synergistic effects that may occur among more than one contaminant (e.g., PAHs + PCB, or Hg + PCB).

3.1(g).1. Uncertainties were identified and discussed within reason. Improvements can be made by discussing the uncertainties associated with using median HQ values as deterministic risk values, and discussing the uncertainties associated with multiple toxins present at levels that exceed benchmark values.

3.1(h).1. The weight of evidence analysis is not very transparent in the document as presented. Please see comment B.2 above for details.

3.1(i).1. Given the data at hand and the description of methods, risk estimates were appropriately derived for reaches of the river where site-specific studies were not conducted.

3.1(j).1 In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local populations of ecological receptors? The conclusion of high risk due to PCBs to invertebrates in the PSA is supported. I believe that the other COCs warrant further attention.

3.2 Amphibian Population Modeling

3.2(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?

Opinion is generally yes. More effort seems to have been exerted on amphibian individual-level and population-level effects than some of the other receptors, and it is not clear why. Why were deformities and sex ratios considered important for amphibians, but not other species? Why were development and maturation used for amphibians, but not other species? Also note, the statement of assessment endpoint in highlighted box (p. 4-1) does not match the formal statement of the assessment endpoint.

3.2(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?

Why were no reference areas used? Why did they not evaluate relationship between number of adults versus number of eggs? The deficiencies in the GE studies were appropriately noted and the studies were generously incorporated into the ERA.

3.2 (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?

This section was appropriate.

3.2 (d) Were the effects metrics that were identified and used appropriate under the evaluation criteria?

There is a need to clarify terminology for frog studies. Controls=frogs from Carolina Biological Supply. Reference=frogs from reference ponds.

Discussion of choice of MATC values is warranted. A sediment MATC of 3mg/kg was set even though EC20 values were much lower than this, especially the EC20 for altered sex ratio. Altered sex ratio should be considered a significant effect since literature studies indicate that transformed "females" are likely not fertile. Further examination of this data is needed.

3.2 (e) Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

I have the same concerns about transparency and ability to determine how statistical procedures were conducted as in previous sections.

The amphibian population modeling exercise was interesting and enlightening. However, there are issues that need to be addressed concerning the parameterization of the model. GE's critique of the model indicated that density-dependent effects can alter reproductive rates of amphibians. The EPA version of the model is set so that the only outcome is extinction of the population. Consideration of density-dependent effects may alleviate this instability in the model and allow for better predictions of population trajectories in reference and impacted areas.

3.2 (f) Was the characterization of risk supported by the available information, and was the characterization appropriate under the evaluation criteria?

The overall characterization is supported and appropriate.

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

Uncertainties were addressed adequately.

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

See comments in section B for WOE analysis.

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

There is high uncertainty characterizing the downstream reaches. Apparently no samplings were conducted downstream of Woods Pond, there is little discussion of habitat suitability, and the risk estimates were based on measured/modeled PCB data.

3.2 (j) In the Panel members' opinion, based upon the information provided in the ERA, does the evaluation support the conclusions regarding risk to local populations of ecological receptors?

Within the PSA, yes. However, further discussion of downstream reaches is needed.

3.3 Fish

3.3(a). In general, studies and analyses were done as described and meet accepted scientific practice. However, it is my opinion that the assessment endpoints were too restrictive for fish in this assessment. Presence of disease and deformities should be used as indicators of population health, but were not apparently used in this way. Studies should also have been done to assess population structure and growth modeling (as with amphibians), but were not. Insufficient data were collected to determine these higher-level effects.

There were some deficiencies in the dose response relationships in the egg injection studies that need to be addressed. Examination of the data in attachment F.7 indicates that there were problems with the Largemouth bass studies. There did seem to be a clear dose response, however, using the rainbow trout model.

3.3(b). Studies conducted by GE to assess the presence of spawning activity and natural reproduction of fish in the PSA were appropriate given the narrow scope of the study design, used accepted scientific methods, and were incorporated appropriately into the ERA. However, no attempt was made to correlate levels of contaminants and these measures, and (in the ERA) there is no analysis of the amount of potential habitat occupied. In addition, there is little discussion of other sources of recruitment (e.g., non-contaminated tributaries or contaminant "cold"-spots in the PSA).

3.3(c). Estimates of exposure and the COC refinement seemed appropriate, but EPA needs to recognize that COC's other that PCBs can cause many of the same effects in fish. A better accounting of the other COC's needs to be present. The Weston "in prep" document should be included in the ERA documentation.

3.3(d).

The effect metrics that were identified and used were appropriate under the evaluation criteria. However, there needs to be discussion of effects metrics that were not identified or used. For example, there appears to be lower rates of natural reproduction of fishes and higher rates of significant deformities in the PSA compared to reference areas. Effect metrics for these areas should have been identified and analyzed. In addition, community structure was not identified as an effect metric, but was used in the assessment.

In addition to identifying appropriate metrics, additional discussion or consideration of the level of the metrics used is merited. When asked why the 50% effect level was used to set toxicity benchmarks, I was told it was non-conservative and it was the "most accurate" estimate of toxicity (EPA answers to Oris written questions). However, in a toxicity frequency distribution, the 50% endpoint is the most precise endpoint estimate of toxicity, not the most accurate. The most precise endpoint provides smaller bounds of error around the endpoint estimate. The most accurate endpoint is determined by how the distribution is modeled. Thus, there seems to be some basic confusion about the choice of the endpoint. In addition, I would argue that multiple endpoints should be used (e.g., 20% and 50% as invertebrates, or the use of slope and center point of the dose response relationship) in the toxicity metric. A paper by Oris and Bailer (Environ. Toxicol. Chem.: 16: 2204-2209. 1997.) describes the rationale for this statement.

3.3(e)

The statistical approaches used in the fish section are not always clear and are not fully explained in the document. I spent most of my review time on the ERA reading the fish section, and the more I read, the less clear the statistics section becomes. Thus it is not possible to evaluate whether statistical techniques were properly applied in all cases since insufficient detail is present to make such an evaluation. Analysis of statistical power, when appropriate, is not provided and queries from panelists concerning power analyses were not adequately addressed. The approach described in Appendix C.2 for handling non-detects would appear to be appropriate, but it is not clear that the procedure was used in all cases. Sections in both fish and invertebrate assessments refer to Appendix C.2, but then unless I missed something, non-detects in the invertebrate assessment were set at 0.5 of the detection limit and were set to the value of the detection limit in the fish assessment. In both assessments then, it appears that fixed-replacement was chosen (with little evidence to support the decision), but two different replacements were used. There seems to be a disconnect between the two assessments and it is not clear how these values were chosen. Were the statistical techniques used clearly described, appropriate, and properly applied for the objectives of the analysis?

3.3(f).

The general characterization of risk is moderately supported by the available information and appears to be within reason, but it is my opinion that potential risk to fish is under estimated. This is because of how the assessment endpoint was defined and evaluated, restricted solely to toxicity testing and very high levels of effect being used to derive benchmarks. The simple presence of fish in the river does not mean the population is healthy (as evidenced by the amphibian modeling exercise). The presence of disease and deformities in the PSA indicates further evaluation is necessary. In addition, because of the way HQ values were used, several instances of risk potential were discounted (further explanation is in 3.3(j) below).

Risk terminology used within the fish section needs to be revised and brought into concordance with other assessments (e.g., wildlife sections). Based on current terminology, any HQ value less than 1 presents no potential risk. HQ values greater than one present varying levels of risk on what seems to be a scale of HQ=1-3 (low risk), HQ= 3-10 (moderate risk), and HQ>10 (high risk). However, this scale is neither formally defined nor justified in the document, and the scale seems to be arbitrary. Based on the methods of developing benchmarks (non-conservative) and the use of extremely high-levels of effect, all HQ values greater than 1 should indicate significant potential for risk. Levels of HQs that approach 1 may also indicate risk potential. For example, there is little possibility that there is a difference in risk potential between an HQ=0.9 and HQ=1.1, yet under the current scenario, the HQ=0.9 is considered "no risk".

Application of HQ values in the risk characterization is wholly deterministic, even though a probabilistic treatment could be used (such as in wildlife assessments). It is strongly suggested that the risk terminology and the risk characterization be conducted using the approach used in the wildlife sections.

3.3(g). Uncertainties were identified and discussed within reason. Improvements can be made by discussing the uncertainties associated with using median HQ values as deterministic risk values, and discussing the uncertainties associated with multiple toxins present at levels that exceed benchmark values.

3.3(h). The weight of evidence analysis is not very transparent in the document as presented. Please see comment B.2 above for details.

3.3(i). Given the data and resources available for these studies, risk estimates for non-site-specific areas appear appropriate, but should carry lower weight in determining risk. Further studies downstream of PSA may be needed to confirm what should be considered preliminary conclusions. Further studies should be conducted to determine whether altered population age-structure is due to lack of harvesting of older fish or lack of recruitment of young fish. No in situ fish toxicity tests were done and all data on fish toxicity, thus, is inferred.

3.3(j). It is my opinion that potential risk to fish in the PSA is underestimated. This is because of how the assessment endpoint was defined and evaluated, restricted solely to toxicity testing and very high levels of effect being used to derive benchmarks. The simple presence of fish in the river does not mean the population is healthy (as evidenced by the amphibian modeling exercise). The presence of disease and deformities in the PSA indicates further evaluation is necessary. In addition, because of the way HQ values were used, several instances of risk potential were discounted. As explained in 3.3(f).2 and 3.3(f).3, the use of deterministic HQ values with an absolute cut-off of HQ<1 presents no risk is problematic in the fish assessment. Figures F.4-6 and F.4-9 indicate that the distribution of HQ values for several fish species span the value of 1, even though the median HQ value was less than one. For example, in Figure F.4-6, only three of the listed median HQ values exceed 1. However, six of the HQ distributions have a 20% or greater chance of exceeding the value of 1 (method for defining "low risk" in wildlife sections). Thus, using the criteria set for fish risk characterization versus wildlife risk characterization, one may make quite different conclusions.

3.4 Insectivorous Birds

3.4(a) As mentioned by one of the public speakers and discussed in the panel, the conceptual model for insectivorous birds needs to be revised or refined. Lumping robins and swallows into one assessment endpoint is not appropriate since they have different food sources. These should either be separate endpoints or they should be used within this endpoint but treated separately through the final risk characterization. I have similar comments on bird assessment endpoints as for fish. The use of individual-based metrics that have been shown to be predictive of higher level effects (e.g., immune function) should be considered and discussed.

3.4(b). There were multiple issues with the GE study that are common to all of the GE studies (e.g., no reference areas, limited study design, too narrow of question asked), but the EPA accounted appropriately for these limitations and incorporated the information appropriately into the ERA.

3.4(c) . More discussion of other COCs is warranted since it appears that data were actually collected on the COCs but were qualitatively discounted in the ERA (Section G.1). Site specific location of nests and local soil, sediment, and food PCB concentrations would help the exposure estimates. Is it possible to use the TDI model to link site specific data to the exposure of birds at specific nesting sites?

3.4(d). Given the statement of the assessment endpoint, the metrics were appropriate. However, there needs to be a discussion of why other endpoints were not used that also would be appropriate. It could be argued that these other endpoints are "individual" level, but many of these endpoints (e.g., physiologically based: gonad weight, enzyme levels, hormone assays, immune function) have been shown to be predictive of long-term effects in birds.

3.4(e). The TDI model was appropriate and well-described. Statistical techniques need to be described in more detail so that it is clear how the data were analyzed throughout the document.

3.4(f). Yes, appropriate.

3.4(g). Uncertainties were appropriately described.

3.4(h). See previous comments on WOE analysis. In addition, the EPA should consider down-weighting the Robin study due to its limitations and deficiencies.

3.4(i). Yes.

3.4(j). Yes, supported.

3.5 Piscivorous Birds

3.5(a) During the GE comments, the use of osprey as an endpoint was criticized. However, based on comments from other public speakers and discussion of the panel, I believe the use of osprey was appropriate. Other endpoints, studies, and analyses were appropriate.

3.5(b). GE study was, again, very limited in scope and suffered from serious limitations. No reference areas were included, no clutch sizes or hatching success were noted, the range of PCB concentrations was too narrow to explore a dose-response, and the sample size was too

small to make any conclusions about the study. This was an admirable attempt at a difficult study, but it does not add a lot of meaningful information to the risk assessment.

3.5(c) . See comments in 3.4. Also, the EPA should consider using exposure estimates as a distribution of concentrations based on the size distribution of prey items, instead of a single point estimate of exposure.

3.5(d). See comments in 3.4

3.5(e). See comments in 3.4

3.5(f). Yes, appropriate.

3.5(g). Uncertainties were appropriately described.

3.5(h). See previous comments on WOE analysis. The weighting used for the kingfisher study were too high. EPA should reconsider weighting assigned to this study.

3.5(i). Yes. However, the uncertainty of using literature derived MATC values should be discussed. Conclusions based on these estimates are not site specific and have high uncertainty.

3.5(j). Within the limits of the available information used in this section, the conclusions are generally supported. Some discussion is needed on the discrepancy between the modeled effects (high risk) and the field studies (low/no risk).

3.6 Piscivorous Mammals

3.6(a) Studies and analysis were appropriate. However, some of the data collected were not used and should be applied to the risk characterization. Specifically, jaw deformities and kit deformities should be accounted for in the risk assessment. These are significant impacts to the organisms and are predictive of long-term effects in the population. Again, I have the same comments concerning the definition of the assessment endpoint as for fish and birds.

3.6(b). The GE studies in this section seem to have the most inherent limitations. Most of these were discussed in the ERA, but in question and answer sessions during the document presentation meeting and the public panel meeting, it became clear that the methodologies used by GE personnel were not all appropriate. These studies should be severely down-weighted or not used at all.

3.6(c) . Some attention should be given to limitations of the feeding study. The MSU study used only goldfish and carp, but mink certainly eat more than these fish. This could have caused a bias in the PCB congeners in the exposure as well as other COCs. Some linkage with the analytical chemistry data is necessary here to alleviate concerns about the dosing regime.

EPA needs to address GE's critique of no evidence of dose response in kit survival.

3.6(d). As mentioned in 3.6(a), there were effects measured, but not used. Several sublethal endpoints such as jaw lesions, could be used here. These provide an additional line of

evidence, are indicative of delayed effects, and are pertinent to population health. Unfortunately, they are not addressed by the assessment endpoint as currently stated in the ERA.

3.6(e). See comments in 3.4. Much of the methodology is simply cited from other reports. The details need to be in the document.

3.6(f). Combining mink and otter (Table I.4-4) in the risk characterization should not be done. They should be treated as separate receptors.

3.6(g). Uncertainties were appropriately described. Analytical uncertainties should be included.

3.6(h). See previous comments on WOE analysis. GE studies should be downweighted due to methodological limitations.

3.6(i). Yes.

3.6(j). Yes, supported..

3.7 Omnivorous/Carnivorous Mammals

3.7(a) Generally appropriate. I question the combination of all of the possible receptors in this feeding category into one receptor. Justification of this combination needs to be made. For example, is it appropriate to combine shrews and fox? It may not be reasonable to separate out all of the different feeding strategies and guilds, but some discussion of how this may add to uncertainty and how the outcome of the risk assessment may be affected is warranted.

3.7(b). As with all GE studies, this one was designed to address a very narrow question and suffered from design and analysis limitations. There were no reference sites, sample areas were limited in number and had high variation in habitat quality. Habitat quality was not addressed. The data were analyzed using each grid as a single replicate, thus there was a sample size of 3 “high” PCB and 3 “low” PCB sites. This is minimal sample size and thus renders statistical analysis nearly powerless. The lack of power allows for the analyst to conclude “no effect” when, in fact, there may be an effect. The study was designed to support a “no effect” conclusion and thus should be severely discounted.

Upon reanalysis, the EPA showed a significant relationship between PCB concentration and shrew survival. The two results are so divergent that it is clear that GE and EPA conducted very different analyses. Upon question at the public panel meeting, the differences were clarified.

In my opinion, the GE analysis was not appropriate for this particular data set. The use of standard probit analysis is not supported by the data quality. GE and EPA techniques are equivalent only if there is a montonic dose response relationship. The probit analysis used by GE cannot account for curvature in data. However, the Bailer and Oris technique used by EPA can model dose responses that include curvature. When a dose response relationship is nonlinear, this will result in different answers between the two methods.

In the case of the shrew data, the dose response was non linear. Use of standard probit model in this case is inappropriate because it violates requirements for a valid probit (e.g., don't use "zero effect" doses, need 3 partial effect levels, monotonic (linear) response. The Bailer&Oris

method can be used to parameterize for any shape of dose response relationship, using a polynomial regression and specifying the appropriate link function (e.g., probit or logit transformation) and error distribution (e.g., binomial for survival or presence/absence). The Bailer&Oris method with a 1st order polynomial (i.e.., a straight line), a probit link, and a binomial error distribution is equivalent to a standard probit regression. The Bailer&Oris method, however, with a 2nd order polynomial will fit a smooth. In this case, the straight line probit would indicate that there was no significant effect within the concentration range tested, but the Bailer&Oris model would.

As a separate issue, the GE analysis used only one data point at each site. If this is the mean of several points, you have thrown-out data and reduced the power of testing for differences between concentrations. The use of individual data points instead of using the means will generate tighter confidence limits around the regression line, the slope value, and the intercept value. This could make the difference between saying that the slope of the line is not different from zero (i.e.., no dose response) versus saying that the slope is different from zero (i.e.., yes, dose response). In other words, using just mean values in the regression could result in an incorrect conclusion of no dose response when one, in fact, existed, regardless of the shape of the dose response (straight line or curved). These two issues (curved dose response and using the mean value at each concentration) compound the situation and make it even easier to incorrectly state that there is no dose response.

3.7(c) . Within the limitations of using literature-derived values, the exposure estimates are appropriate.

3.7(d). Yes, appropriate.

3.7(e). See comments in 3.4

3.7(f). Yes, appropriate.

3.7(g). Uncertainties were appropriately described.

3.7(h). See previous comments on WOE analysis.

3.7(i). Yes.

3.7(j). Yes, supported..

3.8 Threatened and Endangered Species

3.8(a) Within the limits of the study designs, the studies were appropriate.

3.8(b). None were performed

3.8(c) . EPA should consider how habitat use may affect the exposure assessment outcomes. Is it realistic to assume utilization by eagles of only Woods Pond? Can the diet of bats realistically be extrapolated from the swallow diet, even though they occupy temporally different niches?

3.8(d). Effects metrics were defined appropriately

3.8(e). See comments in 3.4

3.8(f). There is some confusion in how the risk was characterized for eagles. Considering that the risk was determined to be "high" for eggs and "low" for adults, how is it possible that the overall risk is "moderate"? It would seem appropriate that the risk should be determined based on the most sensitive life stage. Loss of fish-eating bird populations due to DDT exposure was because of impacts in eggs and not directly on adults, but the risks were high and the impacts were severe. It makes no sense to take the apparent average of the two risk characterizations between eggs and adult.

3.8(g). Uncertainties were appropriately described, but more discussion is needed in terms of analytical uncertainties

3.8(h). See previous comments on WOE analysis.

3.8(i). Yes.

3.8(j). Given the limitations in the studies and the high uncertainties associated with the extrapolations used, the conclusions are supported. However, the EPA needs to clearly articulate the high uncertainty here.

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?

It will be necessary and important for the document to address the issue of whether the presentation can be used to support a remediation decision as well as to set clean-up goals. EPA guidance suggests that a good risk assessment be used to do this. I understand that this particular ERA is different than others since it was defined by the consent decree, but this should be highlighted and discussed. It would be useful for a description of the process by which clean-up goals and remediation decisions will be made.

In addition, further clarification on the extent of risk into CT should be addressed. It appears to me that further study downstream of Woods Pond and into CT is needed, and this risk assessment barely addressed the downstream issues.

I think it would be appropriate to describe the issues surrounding the concerns of public commenters for the CT reaches. Namely the extent of the flood plain and the involvement of Native American tribes in the process. It was clear from the questions I asked of EPA that CT agencies were given full opportunity to participate, but did not participate at a level that may have been appropriate. This may have been a political decision on their part, but the attempt to involve them in the process should be noted without being judgmental.

There are items of revision and clarification that need to be made in the summary discussions and conclusions. For example, the word "catastrophic" is used only in the fish section and not any of the others. It could be argued that none of the receptors are experiencing "catastrophic" failure, so why only point this out for fish? Do we need catastrophic failure in fish populations before we are concerned? It is implicit throughout the document that a HQ>1 is an unacceptable risk; however, this needs to be explicit throughout. The application of risk

characterization needs to be clearly defined and applied consistently throughout the document, but currently it is not.

5. To the best of the Panel's knowledge, is there other pertinent information available that was not considered in the ERA? Is 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.

Consideration EPA guidance on use of societal value for fish receptors should be included.