Peer Review of the Ecological Risk Assessment for General Electric (GE)/ Housatonic River Site, Rest Of River

PREPARED FOR: SRA International Inc.

PREPARED BY: Brad Sample/ CH2M HILL

DATE: February 16, 2004

Attached are my peer review comments on the US EPAs series of documents entitled: "Ecological Risk Assessment for the General Electric (GE)/ Houstatonic River Site, Rest of River". My review has been conducted in accordance with the charge for the ecological risk assessment peer review, with each specific question individually addressed. I scanned all materials provided, but, as my expertise is in wildlife toxicology and risk assessment, I focussed my attention (and the bulk of my comments) on these sections. Comments on the sections dealing with benthic invertebrates, fish, and amphibians are more limited. My review focussed on the materials presented in or otherwise associated with the appendices. I did little more than a cursory review of the associated chapters in the main text.

Both EPA and GE should be commended for the extensive time and effort they have clearly invested in this assessment. In addition to possessing one of the most extensive ecological risk assessment datasets I have seen, this report, and voluminous associated supplementary material, provide an example of the application of many state-of-the-art methods in ecological risk assessment.

Before addressing the charge questions, I have a couple of general presentation-related comments. Although the report is very well written, given the complex data and highly technical material, and because of the massive volume of information evaluated and presented, it can be very difficult to follow. The presentation style used, in which there was a main text that presented a somewhat abbreviated version of the analyses, with the full analyses presented in an appendix, added unnecessary length to the document. I would recommend that the materials in Chapters 3 through 5 and 7 through 11 be replaced by the more detailed discussions presented in the appendices for each respective assessment endpoint.

In addition, because the data used in the assessment are presented in various volumes and appendices, it was very difficult to locate specific data used for each assessment endpoint. To help correct this, I would recommend that all data (i.e., abiotic and biotic chemical concentration data used to estimate exposure, including maps displaying where the data were collected in relation to each other; plus results of field and laboratory studies conducted by both EPA and GE) used for the evaluation of a particular assessment endpoint be specifically reported in association with that evaluation. Although this could result in

the repeated presentation of some data, I think the increased transparency that would result would offset any additional length.

Specific responses to the peer-review charge questions are presented below.

Peer Review Charge Questions

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?

- In terms of characterization of contamination in abiotic media, Reaches 5 and 6 in the Houstonic River watershed appear to be very well characterized. Abiotic characterization of downstream reaches is much less extensive. In contrast to the abiotic media, characterization of contaminants in all biota types is not as robust. The number of biota samples and the spatial distribution of these samples is more limited. Further, because maps presenting biota sampling locations to concentrations in abiotic media are not presented, it is difficult to determine whether the biota chemical data adequately represent the full range of concentrations addressed by the abiotic data. It is essential that biota concentration data represent the locations with maximum COPEC concentrations in abiotic media.

- Section 2.5 - Although I recognize that PCBs are the primary COPC for the site, details on the distribution of other COPCs should also be presented. If not in this section then at least in an appendix. At a minimum, dioxins/furans need to be included in this section as they are the second-most dominant COPC. It is important to know how TEQs vary in relation to PCBs. Another issue - although graphical presentation of data is very useful, tabular presentation of the data (summary statistics by reach) plus information about the datasets on which they are based (when where they collected, by whom, sampling methods used, sample handling proceedures, etc.) should be provided. It is very difficult to track the data sources and summaries for this assessment.

- Figures 2.5-5 through 2.5-11: These figures are very useful for displaying the spatial distribution of PCBs in floodplain soils. Similar figures that displayed PCBs in sediments and TEQs in floodplain soils and sediments would aid the understanding of the spatial context of contamination. An additional enhancement would be to include notations as to where biological sampling was conducted - key maps would display locations where biota sampling for tissue analyses was performed. Additional maps that showed locations where field studies were conducted (i.e., swallow nest boxes, kingfisher burrows, robin nests, small mammal population surveys, etc.) in relation to soil/sediment concentrations would greatly enhance the utility of the report.

- Much more detailed information on the biota chemistry data are needed. Details on where all samples were collected cannot be readily located in the report. Details on how all samples were collected and processed, which analytes were measured in each, and whether samples of abiotic media were collected at the same locations are not readily apparent in the report. Because this report relies heavily on these data, much more detailed information on the chemical characterization of biota is needed.

- Report is not transparent as to which data were used and where they reside in the report. A road map or data inventory that tracks all data that contributed to the decision for each assessment endpoint is needed. The data inventory should include where the data may be found in the report, what they are comprised of (i.e., field analytical chemistry data versus field survey data versus laboratory bioassay data), and the spatial location from which they were obtained and therefore represent. Inclusion of something like this in the beginning of each section would allow readers to develop a better understanding of the depth, robustness, spatial representativeness of the data upon which the risk conclusions for each endpoint are based.

- Characterization of the Connecticut section of the river does not appear adequate – biota sampling and sediment sampling much more limited. I recommend that EPA consider more sampling in this area, especially in depositional areas behind the multiple dams that occur here.

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?

Screening of COPCs

- This screening evaluation is among the most extensive and detailed I have seen. This evaluation would pass as a baseline for other sites. An extensive effort was exerted to reduce the list of analytes and focus on primary issues. EPA looked at all data (point by point) using reasonable toxicity values. It should be noted however that this is not a conservative screen.

- Soil PRGs for wildlife from Efroymson et al. are based on LOAELs and are not really suitable as screening values. No real change needed - just needs to be noted that screening that results from these values is not conservative.

- Background screening was performed for organics. This is generally not performed. Organics are almost exclusively anthropogenic, so there should not be any present in background. Background screens are generally only conducted for inorganics that, because they are derived from underlying geology, may occur naturally. Suggest adding brief text explaining that the screening of organics is non-standard and discussing why it was conducted.

- Appendix B, Page B-22, Section B.4.3: What is the basis for the 10% criteria for exceedences under the 'unlikely' category?

Assessment and Measurement Endpoints

- As currently presented, the assessment endpoints are poorly defined. This issue carries through the whole report and affects how data are used and interpreted. Based on EPA guidance (EPA 1998), a complete definition of assessment endpoints has two components, 1) entities to be protected and 2) attributes that may be affected or at risk. Implicit in this definition is some level of ecological organization (i.e., individual-, population-, community-, or ecosystem-level) for which protection of the assessment endpoint is sought. None of these components are clearly defined for any of the assessment endpoints described in Section 2.8 nor in Table 2.8-1. For individual entities

(identified as receptors in Table 2.8-1; e.g., benthic invertebrates, amphibians, fish, etc.) , the attributes (identified as assessment endpoints in Table 2.8-1) more than one level of ecological organization are implied.

For example, the 'assessment endpoints' for benthic invertebrates are community structure, survival, growth, and reproduction. Whereas community structure is a community- or ecosystem-level entity, survival, growth, and reproduction are individual- or population-level entities. Different data (measurement endpoints), different analyses, and potentially different conclusions may be drawn depending on the level of ecological organization evaluated. I would recommend that the level of ecological organization considered for each assessment endpoint be explicitly stated to ensure that it is correctly matched with appropriate measurement endpoints and so that suitable conclusions can be drawn.

- Note that the Sediment Quality Triad is not really a measurement endpoint in and of itself, rather it is a risk characterization approach. Each of the components of the triad are measurement endpoints though. Suggest removing it from table and discuss in text only.

- Swallows and robins and shrews and foxes do not represent the same ecological exposures. Whereas swallows are aerial insectivores, robins are ground insectivores. Similarly, whereas shrews are ground insectivores with small home ranges, foxes are wide-ranging predators consuming a diverse diet. Pooling these receptors together oversimplifies and misrepresents the exposure pathways being evaluated. I would recommend that these, and the other wildlife species (piscivorous birds and mammals, special-status species), each be evaluated as separate assessment endpoints. Conclusions about potential risks to different ecological guilds can then be made following the weight-of-evidence for each individual species. This is more appropriate given that the level of assessment is at the local population level, not the community level.

- Volume 1 page 2-66 lines 12-17: This paragraph is of great importance. It is the first time that the level of ecological organization for which the assessment is being conducted is discussed. It also is the first (only?) time at which the disjunction between the level of organization associated with the available data and the level of organization for the assessment endpoint is discussed. This text needs to be greatly expanded and integrated (referenced?) into the discussion for each assessment endpoint.

3) For each of the eight assessment endpoints evaluated in the era (listed in Attachment B), address the following questions:

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?

- My background in the evaluation of potential effects on benthics invertebrates is limited. However, the studies, methods, and approaches used by EPA appear to be suitable and appropriate.

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?

- Although GE did not generate any data for benthic invertebrates, they did perform additional analyses of the community data in relation to PCB concentrations. These multiple regression analyses suggest that PCBs may account for 1% – 9% of total variation in community metrics (data presented in January meeting – analyses do not appear to be represented in GEs comments dated September 29, 2003). From the limited description provided at the January meeting, these analyses appear to be suitable and appropriate. I recommend that GE provide the complete analyses and results to EPA for consideration and inclusion in the ERA.

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?

- The sediment concentration data were highly variable over very small spatial scales. This posed a significant difficulty in associating field and laboratory data to appropriate exposure data. EPA generally did a good job of describing the variability and attempting to account for it in their analysis. There were however some problems with the strength of the association of exposure data to effects data that add uncertainty and somewhat reduce the overall strength of the analyses as presented.

- It appears that the exposure data used to determine the dose-response relationship for tPCB was the median concentration observed among all sediment samples collected in 1999. I believe that this is an error. The bioassay results relate best to the sediment concentrations to which the test animals were exposed – this should be the 'most synoptic' sample concentrations. The dose-response-relationships used for the laboratory bioassays, and all subsequent analyses that use these results, should be revised to take this change into account.

- From the discussion at the January meeting, it appears that sampling of benthic invertebrates differed between the shallower upstream areas and the deeper downstream areas. This difference is not immediately apparent in the ERA report. Whereas in the upstream areas, invertebrates were collected in near proximity to sediment samples, this was not possible in the downstream areas. Sediments in the downstream areas were collected by boat. Benthic invertebrate samples were collected from locations near the shore – some 10-20 m distant from the sediment samples. This spatial discontinuity needs to be made clear in the report and described in the uncertainties section. The key question that needs to be looked into is can a gradient exposure be inferred based on these data? Given the high spatial variability, these downstream data may not be adequate.

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

- There are some issues with the effects metrics as presented that need to be addressed:

1) As described in 1.c - site-specific effect levels were based on the median sediment concentrations. I believe that this was in appropriate. The 'most synoptic' values should be

used instead. This will result in a change in the effects levels and will need to be carried through the rest of the section.

2) The interpretation of the bioassays as presented includes the use of discontinuous dose-response relationships (less severe effects observed among some treatments with higher doses). For example, this is the case for *Hyalella* survival in Figure 3.3-1 based on median PCB concentrations. These relationships call into question the suitability and validity of the dose-response relationship. It may be an indication that other factors or COCs may be contributing. (Note that this if the exposure is based on the most synoptic data, this discontinuous dose-response is not an issue…at least in Fig 3.3-1). If discontinuous dose response relationships are observed, the reasons why this occurred should be evaluated and discussed. Additionally, if the anomalous dose-response values cannot be explain such that they can be excluded from the set of values, these bioassay results should not be used to calculate toxicity values.

3) I am not sure that daphnids are suitable for evaluating sediment toxicity. Although they are suitable to look at bioavailability of PCBs, they are only exposed via water and not through ingesting sediment. These data, while useful do not contribute much to the sediment assessment – they may be more suitable for fish assessment.

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

- In general it appears that the statistical analyses were correct. However, adequate detail in the description of the methods and exact data used are lacking. Consequently a definitive conclusion cannot be made. This is a general global statement for all statistical analyses throughout the report. There is insufficient transparency in relation to the exact methods used and whether the data met the assumptions for the method. I would recommend including a brief description in each section about the statistical methods employed, and possibly include a discussion in Appendix C in which detailed descriptions of the statistical methods, their application, assumptions, and limitations, are provided.

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

- The risk characterization appears to be generally appropriate with two exceptions – one being the derivation of the MATC and two being the application and interpretation of this value.

MATC derivation:

- MATCs were calculated by combining both acute and chronic bioassay results. To allow greater resolution in the nature and magnitude of potential effects, these data should not be combined. Rather, separate MATCs should be calculated based on results from acute and chronic bioassay results.

- MATCs were calculated based on multiple measurements of the same type of effect for the same test species. The effect of this is to potentially over-represent effect types and receptors in the derivation of the MATC. I recommend using only a single value for each effect type for each test species per MATC (i.e., use only one *H. azteca* survival value).

This issue will be partially addressed by differentiating/calculating acute and chronic MATCs.

- MATCs were calculated based on sediment-associated test species and water-column test species. I think that this is inappropriate – given questions concerning exposure for the water-column test species, only sediment-associated species should be used.
- To ensure a defendable dose response relationship, test results should be based only on the most synoptic data.
- Any toxicity value for which a discontinuous dose-response relationship was observed should be excluded.
- It appears that the MATCs were calculated based on the lowest six toxicity test values. Justification for this is not provided. I would recommend that the MATC be calculated based on all acceptable toxicity values.
- It is possible that a derived MATC may produce a value that is lower than what is observed at reference locations. If this occurs, the MATC should be truncated to the reference concentration.
- Based on the comments above, a set of rules for deriving MATCs can be outlined: 1) separate acute and chronic 2) do not use multiple endpoints for the same receptor, 3) use only most synoptic data, 4) only use those tests that displayed a dose response, 5) use only sediment-relevant receptors 6) do not restrict data to the lowest 6 values – use all. 7) if MATC is equal or lower than concentration at reference – truncate at the reference concentration.

MATC (and other sediment quality value) application:

- Once calculated, MATCs were applied to all sediment samples within the PSA. This is good. However, the results of this analyses were presented as mean (range) HQs for each area evaluated. This misrepresents potential risks. HQs are basically a binary measure – exposure exceeds or it doesn't. Magnitude of exceedance is not directly interpretable – interpretation is a function of what the toxicity measure represents. I think a better way to present these results would be to tabulate the frequency of exceedance. The greater the number of samples that exceed the MATC (or other toxicity value) the greater the risk. This approach can be refined by using two toxicity values (say an acute MATC and a chronic MATC). Screening results can then be viewed categorically – frequency of samples that do not exceed anything (no risk), frequency that exceed only the chronic value (moderate risk), frequency that exceed both acute and chronic values (high risk). This point by point analyses also would allow for the identification of hotspots and could provide spatial direction for future remedial actions.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination

is a very significant issue, that while mentioned, is not adequately discussed. Additional discussion of the potential interaction between the various contaminants in sediment on benthic invertebrates is needed.

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

- The weight-of-evidence (WoE) analyses is generally appropriate but is not adequately transparent.
- The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

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

- Risk estimates for downstream area are based primarily on the quality and defensibility of the MATCs and their appropriate application. Comments on MATC derivation and application from 1.f apply here too.

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?

In my opinion the conclusions drawn from this section are probably correct. Definitive conclusions cannot be made however until reworking of the data, as suggested above, is complete.

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?

- The field and laboratory studies conducted by EPA were suitable and appropriate and based on accepted scientific practices. The laboratory-based leopard frog study was only partially successful – laboratory fertilization of eggmasses derived from field collected frogs did not work. This reduces the strength of the data. Additionally, the use of reference frogs from VT adds some uncertainty, but because these frogs were collected from within the region and same time of year, I believe that they are suitable and appropriate.

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?

- GE conducted two amphibian studies – one field survey of occurrence of leopard frog eggmasses in relation to PCB concentrations and one in-situ study of the interaction between PCBs and population density among wood frogs.

- Both GE amphibian studies suffered from several statistical and design issues – selection of treatment groups following data collection (leopard frog study), small sample sizes (both studies), no references or controls (wood frog study), etc. As a consequence, their utility in addressing the risk-related questions was reduced. Given these issues, I feel that data from these studies was used and incorporated into the ERA appropriately.

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?

- The exposure evaluation appears to be suitable and appropriate.

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

- The effects data appear to be suitable and appropriate.

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

- The statistic analyses appears to be generally suitable and appropriate, however transparency in approach is lacking as discussed in comment 1.e (above).

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

- The risk characterization for the field and laboratory studies appears to be supported by the available information and appropriate under the evaluation criteria.

- The risk characterization for the tissue based-MATCs is not adequately described or presented. It is not clear from the text in Section E.4.5, nor from Table E.4-2 or Figures E.4-1 through E.4-3 upon which tissue concentration value the HQs were based. Were HQs calculated based on the mean or median tissue concentrations? This needs to be

made clear. In addition, if HQs are calculated based on a summary statistic (or are done on a point by point basis, then averaged) this is inappropriate. HQs should be evaluated on a point by point basis and then presented as the frequency of samples that exceeded.

- Although a MATC was derived for sediment, it does not appear to have been used to provide a detailed spatial evaluation for the PSA. Such an analysis would be beneficial in identifying potential hotspots and for describing the spatial distribution of exposure and risk.

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 significant uncertainties appear to have been identified and evaluated.

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

- The weight-of-evidence (WoE) analyses is generally appropriate but is not adequately transparent.

- The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

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

- The same issues identified for the development and application of MATCs for benthic invertebrates (1.f, above) apply to the MATCs for amphibians. Use of the same decision process for derivation of the MATC and point by point application as for benthic invertebrates is recommended.

- It appears that data for PCB concentrations in downstream vernal pools are lacking. As these are the habitats in which amphibians are most likely to be exposed and affected, the lack of vernal pool data from downstream reduces strength of downstream analyses. This lack of data and its effects should be noted.

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?

- Based on the information provided in the ERA, the risk conclusions are probably accurate and appropriate.

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?

- The EPA studies appear to be suitable and appropriate.

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?

- GE study appears to have been well designed and conducted. The results of this study however do not appear to have been correlated to the PCB concentrations, so not dose-response association can be made. Data indicate that the population is skewed toward older fish – this could indicate adverse effects on recruitment. Overall, the study was correctly used by EPA.

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?

- The exposure evaluation appears to be suitable and appropriate. Tissue concentration data provide the most accurate measure of exposure. Use of sediment data to evaluate exposure for PAHs is suitable due to the metabolism of these compounds.

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

- The effects data generally appear suitable and appropriate – However, as this is out of my area of expertise, I need to defer to the other panel members on this topic.

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

- The statistic analyses appears to be generally suitable and appropriate, however transparency in approach is lacking as discussed in comment 1.e (above).

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

- The risk characterization as presented in the ERA is probably correct. I do believe that more analyses and discussion of the observation of skewed population demographics (great number of old individuals) should be provided.

- I was pleased to see the point by point evaluation of fish concentrations Figures F.4-1 through F.4-5, plus F.4-7 and F.4-8.

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.

- The discussion of uncertainties is reasonably adequate. A couple of topics are missing that should be included. These include the need to explicitly discuss how PCB and TEQ concentrations vary seasonally in relation to lipid content of fish and how this may affect temporal and season exposure and potentially risk. In addition fish are simultaneously exposed to multiple contaminants. The interactive effects of exposure to multiple toxins should be discussed.

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

- The weight-of-evidence (WoE) analyses is generally appropriate but is not adequately transparent.

- The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

- Weighting for field studies appears low – rated as undetermined for evidence of harm – need to explain how overall weightings were derived

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

- The effects data generally appear suitable and appropriate – However, as this is out of my area of expertise, I need to defer to the other panel members on this topic.

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?

- Based on the information provided in the ERA, the risk conclusions are probably accurate and appropriate.

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?

- The EPA studies and analyses were suitable and appropriate. The 3-year swallow study was very comprehensive, used a good experimental design, and produced fairly robust data. Analyses and conclusions appear to be appropriate. The rather limited amount of effects observed to swallows from PCBs is consistent with the lower comparative sensitivity of tree swallows to PCBs than some other birds. This lower sensitivity of swallows should be acknowledged and discussed as results from this species are being used to evaluate effects for other avian insectivores.

- I think that the analyses would have benefited from the evaluation of another species that is more highly exposed to invertebrates in floodplain soils, in particular, the American woodcock. Although the secretive nature of this species would have made collection of site-specific reproduction data problematic, modeled exposure estimates using measured PCB and TEQ concentrations in food would have provided a good indication of exposure and effects.

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 GE robin study suffers from multiple design and analysis issues. As a consequence, it's conclusions are not strongly supported by the data. First, the authors treat the nests as replicates to test for a treatment effect. In actuality, they are testing a location effect. Whereas all exposed nests were within the 10 year floodplain, all reference nests were over 300 m outside the floodplain. Although PCBs may differentiate between these groups, so might habitat differences (which were not quantitatively measured). Because the 'treatment' groups are spatially distinct and are not interspersed, the study is an

example of pseudo-replication (see Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Mono. 54:187-211). Sample size is 1 treatment and 1 reference, with multiple subsamples in each. Hypothesis testing analyses are therefore not appropriate. A more suitable analyses would be to perform regression analyses to see if there was an association between PCBs in eggs and nest success parameters.

- A potential bias in estimation of robin exposure via eggs exists because sampling was restricted to only viable eggs. Non-viable eggs may have been so due to elevated PCB/TEQ concentrations. Preferential selection of only viable eggs may have resulted in an underestimation of egg-mediated exposure to nestling robins. This potential bias was not discussed in either the GE study nor the ERA. At a minimum, a discussion should be included in the uncertainties section of the ERA.

- Another limitation of the study was that it only evaluated data from a single year. Nest success can vary dramatically from year to year. Some analyses of how ambient conditions (i.e., weather and other environmental conditions) for 2001 related to other years would strengthen the argument that uncommon events did not occur that may have biased the data. Similarly, if detailed quantitative habitat data were available for each nest, these could be used to help increase confidence that the results represent a treatment effect (of lack of it) and not a habitat effect.

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?

- Although the estimation of exposure and effects was generally appropriate, there were several areas that require further evaluation:

1) It was good to see the development and application of the microexposure model for swallow nestlings. It provides a useful mechanism to evaluate exposure and potential effects for what is clearly the most sensitive life-stage. However, I think that the TDI model should also have been run for adult swallows. This is for two reasons: 1)exposure and risks for adults differ from what would be expected for nestlings; and 2) the TDI modeling approach was used for all other wildlife receptors. To be consistent with, and allow comparisons to other wildlife receptors, the TDI model should also be run for adult swallows.

2) Appendix G, Page G-16, Lines 17-27. It appears that the food ingestion rate for nestling swallows was estimated using the allometric model for adult birds from Nagy (1999). The uncertainty or appropriateness of this model to represent nestling food ingestion should be discussed. Nichols et al. (1995) developed a model for nestling food ingestion in their swallow model. At a minimum, food ingestion rates estimated by both approaches should be compared.

3) The TDI modeling effort for robins relied exclusively on directly measured concentrations of PCBs and TEQ in earthworms. Similarly, the microexposure model relied exclusively on PCB and TEQ concentrations in stomach contents. This approach will give good estimates for all locations from which these site-specific biota data were available. However it does not address exposure and risk in areas from which site-

specific biota data were unavailable, nor does it allow exposure and risk levels to be associated with specific PCB and TEQ concentrations in floodplain soils or sediments (information extremely useful to make remedial decisions). No effort was made to model bioaccumulation into biota and estimate exposure at all locations, with suitable, species-specific habitat, from which abiotic data were available. Given the large size of the site, the heterogeneous nature of contamination, and the variable movement patterns of resident wildlife, the current approach may not adequately represent the exposure that local populations of wildlife may experience at the site. It is recognized that modeled concentrations in biota increase uncertainty. However, this would be balanced by the increased spatial resolution in exposure estimates. I do not recommend replacing the current approach. Rather, exposure based on COPC concentrations modeled in biota should be used to provide a spatially more complete estimate. The results from both can be compared as part of the risk characterization.

4) The TDI and microexposure modeling efforts relied exclusively on directly measured concentrations of COPCs in dietary items (either earthworms for robins, or stomach contents for swallows). The results of these analyses, derived from data representing a comparatively limited spatial area, were then extrapolated to represent the whole PSA. For this approach to be valid, it is important that the areas from which biota data were collected represent the range of possible COPC concentrations in abiotic media, in particular it is essential that the highest concentrations be represented. It is not apparent that any analyses were conducted to indicate that the available site-specific biota adequately represented the full range of possible COPC exposures (including maximum concentrations).

5) References for sources of parameter values in Tables G.2-2 and G.2-3 and Tables G.2-24 and G.2-25 are lacking and should be provided.

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

- The effects data appear to be suitable and appropriate. EPA has conducted a very thorough review of the available toxicity data and has made a very strong effort to adequately represent the range of potential effects. One aspect lacking from the ERA was an evaluation of egg-based effects data. Site-specific egg-based exposure data were available for robins (and could potentially have been modeled for swallows). Egg-based PCB effects data are available in the literature (Hoffman et al. 1996, "PCBs and dioxins in birds." In *Environmental Contaminants in Wildlife, Interpreting Tissue Concentrations*. W. Nelson Beyer, Gary H. Heinz, and Amy W. Redmon-Norwood, eds. Lewis Publishers, Boca Raton.).

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

- Similar issue with transparency of statistical analyses as discussed for earlier sections, apply to this section too.

- It is not immediately clear in this section, nor in the subsequent sections and appendices in which Monte Carlo simulation of wildlife exposure is discussed, as to how chemical concentration data were included in the exposure models. However, at the December

meeting, Dwayne Moore stated that chemical concentrations were included in the Monte Carlo analyses as a point value (the 95% upper confidence limit on the arithmetic mean [UCL95]). I am concerned that this approach does not adequately nor accurately represent the variation and uncertainty in exposure. The exposure distribution that is generated only represents the uncertainty and variability about the distributions for the life-history based parameters (body weight, diet composition, food ingestion, etc.). Variability and uncertainty associated with PCB/TEQ concentrations are not represented. Depending on how large the sample size is and the variance associated with the data, use of the UCL95 may over represent exposure. Although the probability bounds analyses captures what should be all possible exposures, it is a very broad brush. The Monte Carlo analyses should provide what is believed to be the reasonable, most-likely exposure distribution based on expected distributions of all parameters.

- Statistical analyses should be provided for data presented in Tables G.19 through G.23 to indicate which sites differed from each other.

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

- Although the risk characterization for insectivorous birds was generally supported by the available information and was appropriate, multiple issues were identified that require attention/revision:

1) The purpose of the risk characterization stated at the end of Section G.4 is to determine whether exposure is 'sufficient to cause adverse effects to individuals inhabiting the PSA'. This purpose should relate back to the assessment endpoint described in the Problem Formulation – I believe that the assessment is supposed to be on populations, not individuals. In addition, some discussion of the spatial aspects of exposure and risk should be included. This is a particularly important issue with a large site such as the Housatonic.

2) To interpret the significance and general applicability of the risk characterization results to the PSA as a whole, it is important to know how the limited number of locations represented by field data or exposure estimates compare to the overall contamination within the PSA as a whole. If the maximum exposure is not represented by these data, then the full range of exposure (and risk) is not adequately represented.

3) Because swallows and robins represent very different exposure regimes, I do not think that they are directly comparable and would recommend that they be separated and evaluated independently.

4) The discussion in Section G.4.1.1 is based solely on the evaluation of exposure and risks to nestling swallows. This should be made clear. Because a TDI model was not run for adult swallows, exposure and risks to adults have not been directly evaluated. The text in this section discusses tree swallows in general although data are only for nestling swallows. Additional discussion is needed to explain how the results for nestlings relate to the larger swallow population as a whole. It should also be noted that risks are estimated based on measured and estimated tissue concentrations – this is different than for robins (next section).

5) Text should be added in Section G.4.1.2 to clarify that the exposure and risk estimates for robins are based on dietary exposure, not tissue concentrations as were swallows.

6) The first sentence of the conclusions portion of Section G.4.2 states that fecundity of swallows does not appear to be affected by contaminants. This directly contradicts the results presented in the preceding section, in which a significant negative relationship between PCBs in eggs and hatching success was observed. This discrepancy needs to be resolved. It might be more appropriate to state that swallows are affected but not severely so.

7) Analyses of variance methods are used to evaluate whether PCBs had a significant effect on nesting success of robins. Use of regression methods, similar to those used for the swallows would probably be more suitable.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination and the degree to which the exposure data actually represent the PSA are very significant issues. These need to be discussed.

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

- As stated previously, the weight-of-evidence (WoE) analyses is not adequately transparent. The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

- The WoE analyses for swallows appears to be appropriate and suited to the available data, except that the evidence of harm from the field study is incorrectly stated as 'No'. This is not consistent with earlier text. On page G-78, Line 15 the statement is made that risks are present but low for swallows. Consequently, I think the table should be revised to state 'Yes' for evidence for harm based on the field study.

- The WoE analyses for robins generally appears to appropriate and suited to the available data, except that I believe that the field study is weighted to high. I think that it should have a weighting no higher than moderate.

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

- A downstream evaluation not done for insectivorous birds. Rather extrapolation to other insectivorous bird species was provided instead. Some explanation should be provided to explain why downstream risks were not evaluated.

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?

- Although the information in the ERA does support the final conclusions, some revision and clarification is needed to indicate that the evaluation is for the local population and not for individuals.

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?

- EPA analyses for piscivorous birds were appropriate and based on accepted practices. Both the kingfisher and the osprey are suitable receptors for the PSA.

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?

- GE should be commended for attempting to collect field data on the resident kingfisher population in the PSA. The methods used are suitable and appropriate for this species. Unfortunately the study suffers from small sample size and limited study duration. As a consequence, utility of conclusions are limited - definitive conclusions about the presence or absence of effects cannot be made. Limitations to the conclusions include: 1) Data were from only a single year, with the representativeness of conditions in the year being unknown. 2) No data from a reference location were available for comparison. 3) low densities resulted in sample sizes being very small. The only conclusion that can definitively be made is that kingfishers do reside in the PSA and they appear to be reproducing. Collection of eggs, nestlings, or stomach contents for residue analyses would have greatly strengthened the data from this study. In the absence of these data, the magnitude of exposure experienced by birds at each nest is uncertain.

- Comparatively high predation rates observed may be suggestive of an effect on nest attentiveness (may also be noise). Note that habitat in area is not of particularly high quality - maximum HSI was only 0.5. Note also that in comparison to nest success (after removal of depredated nests) from other areas, nestling survival from Housatonic is one bird fewer per nest - not statistically significant, but considering the small sample size and limited time duration, this result does raise questions.

- In the WoE EPA considers the results of this study to be indicative of no evidence of harm. Given the limitations of the study, I believe it is more appropriately considered to be inconclusive (i.e., undetermined) in terms of evidence of harm.

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?

- This comment applies to Appendices H, I, and K and relates to fish size in exposure analyses for piscivores. Piscivorous wildlife generally select fish opportunistically based on size. Exposure for kingfisher, osprey, bittern, mink and otter were all based on the UCL95 concentrations in fish that fell within the appropriate size classes. Although the largest size fish in the size range consumed by each species are likely to have the highest

PCB concentrations, it is unlikely that the diet of a piscivore will consist extensively of the largest fish they can consume. Rather, they will consume fish that tend toward the center of their respective size range (these will have lower PCB concentrations). Using the UCL95 for all fish within the size range used by a species may overestimate exposure. [Another issue is that we do not know what the size distribution for fish within each range was - was the distribution even over sizes or biased? This will also affect the accuracy of the exposure estimate] An alternative would be to segregate fish into size classes, calculate UCL95s by size class, then incorporate the appropriate size class concentrations into the exposure estimate for each species.

- Habitat used by osprey. It appears that this species was assumed to be exposed to PCB concentrations in fish from throughout the PSA. This does not seem appropriate. The upper sections of the PSA are generally too overgrown and shallow to allow significant use by the large piscivorous birds. The exposure estimation would be more appropriate if it weighted the exposure to be higher in the southern areas in the vicinity of Woods Pond and lower for the upper areas.

- References for sources of parameter values in Tables H.2-9 and H.2-10 are lacking and should be provided.

- It is not clear from where the fish (and crayfish) used to estimate exposure to kingfisher and osprey were collected. Reach 5 is a very large area. To understand how representative the exposure data are, some reference to the spatial locations addressed should be provided. Maps displaying sampling locations (an associated PCB concentrations in sediment) for biota used in the piscivore exposure would help.

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

- The effects data appear to be suitable and appropriate. EPA has conducted a very thorough review of the available toxicity data and has made a very strong effort to adequately represent the range of potential effects.

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

- Similar issue with transparency of statistical analyses as discussed for earlier sections, apply to this section too.

- As in the insectivorous bird section (comment 4.e above), it is not immediately clear how chemical concentration data were included in the exposure models. However, at the December meeting, Dwayne Moore stated that chemical concentrations were included in the Monte Carlo analyses as a point value (the 95% upper confidence limit on the arithmetic mean [UCL95]). I am concerned that this approach does not adequately nor accurately represent the variation and uncertainty in exposure. The exposure distribution that is generated only represents the uncertainty and variability about the distributions for the life-history based parameters (body weight, diet composition, food ingestion, etc.). Variability and uncertainty associated with PCB/TEQ concentrations are not represented. Depending on how large the sample size is and the variance associated with the data, use of the UCL95 may over represent exposure. Although the probability bounds analyses captures what should be all possible

exposures, it is a very broad brush. The Monte Carlo analyses should provide what is believed to be the reasonable, most-likely exposure distribution based on expected distributions of all parameters.

- Statistical analyses should be provided for data presented in Tables H.2-4 and H.2-5 plus Tables H.2-12 and H.2-13 to indicate which sites differed from each other.

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

- The risk characterization appears to be generally.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination and the degree to which the exposure data actually represent the PSA are very significant issues. These need to be discussed.

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

- The weight-of-evidence (WoE) analyses is generally appropriate but is not adequately transparent.

- The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

- Weighting for field study appears high – given the limitations of the study, I would rank it as moderate at best. In addition, as stated above, I believe it is more appropriately considered to be inconclusive (i.e., undetermined) in terms of evidence of harm.

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

- A downstream evaluation not done for piscivorous birds. Rather extrapolation to other piscivorous bird species was provided in Appendix K instead. Some explanation should be provided to explain why downstream risks were not evaluated.

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?

- Although the information in the ERA does support the final conclusions, some revision and clarification is needed to indicate that the evaluation is for the local population and not for individuals.

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 EPA studies appear to be suitable and appropriate.

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?

- This short-tailed shrew study was suitable and used appropriate methods. However, there are a number of study design issues that limit the utility of its results. The study suffers from the lack of a reference location, small sample sizes, and limited study duration. Habitat quality also appears to be a confounding factor that was not adequately evaluated. Because the data were from only a single year - the representativeness of conditions in the year are unknown. No data from a reference locations were available for comparison. Sample sizes were very small. Vegetation cover was heterogeneous among and within sampling areas. Statistics showed that population parameters differed significantly by area - strongly suggesting a confounding habitat effect - Addition analyses using detailed habitat data may be useful to tease out the habitat effects from contaminant effects. Given the high variability known to exist for shrew populations over years, seasons, and habitats, etc., the fact that the study looked only at one year, had no reference area, and had only 6 locations seriously limits the strength of any conclusions.

- Conflicting conclusions based on the GE and EPA analyses of data from this study need to be better explained. At the January meeting, GE stated that they could not recreate the statistical results produced by EPA. Extensive discussion at the meeting indicated that the analyses conducted by EPA and the re-analyses conducted by GE differed in how they handled the data and the exact statistical methods used:

    - GE/Boonstra – treated grids as reps - did not weight grids – did probit transformation – eliminated grid 3 due to presence of single male – not included in analyses for males, included for male and female.

    - EPA – methodology in Bailor and Oris – used GliM models with – probit transformed link function – each grid is weighted by total number of organisms in the treatment.

    I cannot comment on the statistical applicability of both approaches. I do think that the Boonstra re-analyses should be included in the ERA and the differences between the two approaches and their results should be discussed.

- In the second paragraph of Section J.4.3.3 it is stated that the spatially weighted PCB concentrations were calculated based on the *arithmetic mean* concentration in soils. For virtually all other analyses in this ERA, the lognormal distribution was assumed for abiotic media. An explanation/justification for the use of the arithmetic mean needs to be provided.

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?

- Although the estimation of exposure and effects was generally appropriate, there were several areas that require further evaluation:

1) The TDI modeling effort for fox and shrews relied exclusively on directly measured concentrations of PCBs and TEQ in earthworms. This approach will give good estimates for all locations from which these site-specific biota data were available. However it does not address exposure and risk in areas from which site-specific biota data were unavailable, nor does it allow exposure and risk levels to be associated with specific PCB and TEQ concentrations in floodplain soils or sediments (information extremely useful to make remedial decisions). No effort was made to model bioaccumulation into biota and estimate exposure at all locations, with suitable, species-specific habitat, from which abiotic data were available. Given the large size of the site, the heterogeneous nature of contamination, and the variable movement patterns of resident wildlife, the current approach may not adequately represent the exposure that local populations of wildlife may experience at the site. It is recognized that modeled concentrations in biota increase uncertainty. However, this would be balanced by the increased spatial resolution in exposure estimates. I do not recommend replacing the current approach. Rather, exposure based on COPC concentrations modeled in biota should be used to provide a spatially more complete estimate. The results from both can be compared as part of the risk characterization.

2) The TDI modeling effort relied exclusively on directly measured concentrations of COPCs in dietary items (either invertebrates for shrews, or small mammals for the fox). The results of these analyses, derived from data representing a comparatively limited spatial area, were then extrapolated to represent the whole PSA. (This is particularly an issue for the shrew – PSA exposure is based on invertebrate concentrations from only three locations. For the fox, the issue is that it is not clear from where the small mammals samples were collected – are they representative of the whole PSA? Do they adequately represent the range of PCB and TEQ concentrations in soils in the PSA? Details on small mammal sampling locations needs to be included in this section.) For this approach to be valid, it is important that the areas from which biota data were collected represent the range of possible COPC concentrations in abiotic media, in particular it is essential that the highest concentrations be represented. It is not apparent that any analyses were conducted to indicate that the available site-specific biota adequately represented the full range of possible COPC exposures (including maximum concentrations).

3) References for sources of parameter values in Tables J.2-2 and J.2-3 and Tables J.2-10 and J.2-11 are lacking and should be provided.

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

- The effects data appear to be suitable and appropriate. EPA has conducted a very thorough review of the available toxicity data and has made a very strong effort to adequately represent the range of potential effects.

- All toxicity data were based on rodents. These data may not taxonomically be the most suitable for fox. As they are from the same order (Carnivora), mustelid data may be more suitable for fox than rodent data. At a minimum, the potential effects of using data

from taxonomically dissimilar species should be included in the uncertainty section (existing text expanded).

- In the last paragraph of Section J.3.2.1 it is stated that the mouse was used as the surrogate for the shrew and the rat was the surrogate for the fox. It is unclear as to why this distinction was made. I'm assuming it was due to body size of the surrogate relative to the shrew or fox. The reasoning for this distinction should be explained.

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

- Similar issue with transparency of statistical analyses as discussed for earlier sections, apply to this section too.
- Similarly, the issues with the probabilistic analyses previously mentioned under other receptors apply for the analyses for shrew and fox.

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

- The risk characterization is suitable and appropriate.
- Organizational note: descriptions of the field studies and their analyses only appear in the Risk Characterization section. As these are really effects data, they should be described in the Effects section, with interpretation and integration of results left in the Risk Characterization section. This anomalous presentation of data also occurs for some other receptors. The organization of the report would benefit if the presentation of the data were moved.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination and the degree to which the exposure data actually represent the PSA are very significant issues. These need to be discussed.

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

- The WoE was conducted for omnivorous/carnivorous mammals as a whole and not separately for shrews and fox. Because the same data were not available for each receptor and because the shrew and fox represent distinct exposure pathways, separate WoE analyses should be performed for each receptor.

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

- A downstream evaluation not done for omnivorous/carnivorous mammals. Rather extrapolation to other omnivorous/carnivorous mammal species was provided instead. Some explanation should be provided to explain why downstream risks were not evaluated.

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?

- Although some clarification and revision of the analyses performed are needed, I believe that the conclusions are probably appropriate and are supported by the available data.
- Additional discussion is also needed to indicate that the evaluation is for the local population and not for individuals..

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?

- The field and laboratory studies conducted by EPA were generally appropriate and acceptable. There however are several issues with both studies that should be addressed.

Mink Bioassay

- The PSA fish incorporated into diet for the test mink were represented by large goldfish and carp. I recognize the need to have adequate biomass to create sufficient test diet to sustain the mink for the study duration. However, due to their size, large goldfish and carp are not fish that mink in the PSA would be consuming to any great degree. In addition, the feeding habits of goldfish and carp may potentially result in a different contaminant mixture than other fish. To address this uncertainty, I would recommend including an evaluation of how the contaminant mixture in the fish used for the diet compares to that for the fish assumed to represent the diet of mink in the PSA. Questions to include are, 1) how does congener mixture vary by fish size (age), and 2) how does congener mixture vary by fish species.
- The test diets were only analyzed for chlorinated compounds – what about PAHs (might not be detected) or metals?.
- DDE was elevated in the PSA fish used in the diet. This is a potential confounding factor that needs to be explained more. Potential effects that this may have had should be discussed. Concentrations of DDE in the final diets should be presented if measured. How do these concentrations relate to what is seen in other studies?
- Table 3 from Bursian et al. does not report PCB or TEQ concentrations – this would be helpful to see in relation the other organochlorines.
- Just as a note, it is unfortunate that the dose range used in the study did not extend to one or possibly two higher doses. These higher doses would have likely produced more severe effects and would have strengthened the overall dose-response relationships.

Field Survey

- Although my impression of the EPA study is that it is appropriate and well conducted, GE raised a valid issue at the January meeting: habitat differed between the PSA and the references areas. Whereas the PSA is a river and it's floodplain, the reference areas were all comprised of ponds and their shorelines. This issue should be directly addressed in the ERA. Potential habitat and methodological differences (transects in the PSA versus

whole shoreline surveys in the reference areas), and how they may have affected the results should be discussed when these data are first presented and in the uncertainties section.

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?

- GE conducted an independent field study to evaluate the presence and abundance of mink and river otter in the PSA. The EPA identified numerous limitations to the GE study (Appendix I, Pages I-62 and I-63). Based on my review of the GE report, I concur with the EPA comments. I think that the GE study greatly overstates the abundance of mink and otter in the PSA, and I think that the study was appropriately incorporated into the ERA.

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?

- As discussed for avian piscivores in comment 5.c, the exposure analyses would be strengthened by increasing the resolution of the fish size-class data in the exposure model.

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

- Although a site-specific mink toxicity study was conducted, the results of the TDI modeling were compared to literature-derived toxicity data. The site-specific mink study by Bursian et al. produced NOAELs and LOAELs for kit survival of 0.169 and 0.414 mg/kg/d for PCBs and 1.69 and 7.67 ng/kg/d for TEQs. In contrast, the literature-derived values were 0.128 and 0.0272 mg/kg/d for PCBs and 3.6 and 36 ng/kg/d for TEQ. The results from the site-specific toxicity study should be used to evaluate exposure and risk rather than the literature-derived data.

- The site-specific mink study indicates that the Housatonic PCB mixture is somewhat less toxic to mink than mixtures observed at other locations. As a consequence, use of literature-based data over-estimates risks at the site. My recommendation is that the site-specific data be used instead if the literature-derived values.

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

- Although the statistical analyses in the mink toxicity study are well described and appropriate, the same transparency issues previously mentioned under other receptors apply for the analyses for mink and otter.

- Similarly, the issues with the probabalistic analyses previously mentioned under other receptors apply for the analyses for mink and otter.

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

- Although the risk characterization was appropriate for the field studies, the results of the TDI modeling should be evaluated based on the site-specific toxicity test results and not on the literature-derived results.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination and the degree to which the exposure data actually represent the PSA are very significant issues. These need to be discussed.

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

- WoE analyses had a couple of inconsistencies that need to be corrected.

  1) The WoE analyses combined both mink and otter. This is inconsistent with other sections. Separate WoE evaluations need to be performed for mink and otter.

  2) TEQs were listed as an independent line of evidence – this is not correct and is inconsistent with other sections.

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

- The downstream risk evaluation for mink and otter appears to be appropriate. However there are several issues that need to be addressed before a definitive conclusion can be made:

1) the downstream analyses is only a screening evaluation and significant uncertainties exist. Text needs to be added clarifying that the results only represent a screen and the uncertainties specifically associated with this evaluation need to be identified.

2) The analysis lacks transparency. The text briefly describes how the MATC was derived and states that exposure was based on fish data collected in the downstream reaches between 1999 and 2002. Only in Figures I.4-15 and I.4-16 is it stated that manipulations of the fish concentration data were conducted to estimate exposure. It appears that various scaling factors were employed to adjust fish concentration data. A detailed explanation and justification of these adjustments needs to be provided in the text. The sources for the scaling factors applied needs to be presented.

3) Risks are broadly displayed in Figures I.4-15 and I.4-16 as reaches of the river where exposure either exceeded the MATC or did not. Locations from where fish samples are available are not presented. I suspect that the available data are not spatially continuous enough to allow such broad conclusions. Locations from where fish samples exist need to be presented in the map and the risk conclusions should be based solely on those areas for which fish data are available.

4) What contaminants were measured in the downstream fish?. Were the analyses restricted to PCBs? As the analyses for Reaches 5 and 6 were based on PCBs and TEQ, for consistency, a statement as to whether TEQ data were available downstream needs to be made. If TEQ data are available, they should be evaluated also.

5) The upstream and downstream exposure analyses differed in how they were conducted – (more complete dietary exposure estimation for Reaches 5 and 6, screening of MATC

against fish concentrations in downstream reaches). For the purpose of comparing relative exposures and risks along the course of the river, additional screening of the fish data (using the MATC approach used for downstream locations) from reaches 5 and 6 should be considered. These data could be presented in this part of the section as histograms representing each reach.

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?

- I believe that the risk conclusions are probably supported by the available data. Revisions of the analyses as suggested above are necessary however to ensure that the appropriate magnitude of risks is documented.

- In addition, some revision and clarification is needed to indicate that the evaluation is for the local population and not for individuals.

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?

- Studies and analyses performed by EPA for T&E species were generally suitable and appropriate.

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?

- No studies were performed by GE for 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?

- Exposure modeling for the T&E species was conducted in the same way as for other wildlife receptors. Consequently many of the issues previously discussed for other wildlife receptors (i.e., reliance on measured versus modeled tissue concentrations; degree to which measured data represent full range of potential exposure in the PSA, etc.) also apply for the T&E species. Other specific issues include:

    - Exposure estimation for Bald Eagles - The fish component of the eagle diet is broken down by feeding guilds (predatory, forage, and bottom feeders). Although this is correct, I believe the size-based approach outlined above for other piscivorous wildlife will provide a more suitable representation of exposure for bald eagles.

    - Use of the egg accumulation model for bald eagles is appropriate given the T&E status of the species. However because uncertainty associated with the model is extremely high, some effort to validate the modeling approach should be performed. This would mean identifying at least one location (preferably more than one) anywhere within the range of the bald eagle for which there exist data on measured concentrations in bald eagle eggs plus associated fish concentration data that can be used to estimate dietary exposure. To allow for the correct interpretation of this modeling approach, some information on the accuracy of the resulting estimate is

needed. In addition, it needs to be noted that bald eagle do not currently nest in the PSA. This analyses simply evaluates potential effects that could occur should individuals start to nest at some point in the future.

- Note – the egg exposure modeling approach used for bald eagles could also have been applied to the other avian receptors that nest in the PSA (validation for these species would also be needed to calibrate he accuracy of the method). This would have added an additional measure to evaluate exposure and effects.
- The diet of the bat was assumed to be adequately represented by the stomach content data from nestling swallows. Although this may be correct, additional analyses and discussion are needed to support this assumption. In particular, comparisons relating diet for swallow (unpubl. Custer data) to that of the bat (literature) should be performed.
- References for sources of specific exposure model parameter values need to be provided in Tables K.2-2 and K.2-3, Tables K.2-10 and K.2-11, and in Tables K.2-18 and K.2-19.

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

- Recognizing the limitations in available toxicity data, the effects metrics selected are suitable and appropriate.

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

- The issues with the probabilistic analyses previously mentioned under other receptors apply for the analyses for T&E species.

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

- The risk characterization is suitable and appropriate.
- Organizational note: descriptions of the field studies and their analyses only appear in the Risk Characterization section. As these are really effects data, they should be described in the Effects section, with interpretation and integration of results left in the Risk Characterization section. This anomalous presentation of data also occurs for some other receptors. The organization of the report would benefit if the presentation of the data were moved.

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.

- Although the uncertainties are generally well captured in the ERA report, the relative importance of various uncertainties is not discussed. Interpretation and understanding of the results would benefit by such an analysis. Spatial heterogeneity of contamination and the degree to which the exposure data actually represent the PSA are very significant issues. These need to be discussed.

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

- The WoE for T&E species was conducted as if each T&E species was an independent line of evidence. This is inappropriate. Because the conceptual models and exposure represented by each species differ, a separate WoE evaluation should be performed for each species.

- As stated previously, the weight-of-evidence (WoE) analyses is not adequately transparent. The WoE analyses use best professional judgement to apply final weightings for the various lines of evidence and their attributes. It is unclear how final weightings were derived. For example, it is not clear why evidence for harm for small footed myotis from PCBs and TEQs and for the bittern for TEQs are listed as undetermined. Some discussion as to how these weights were derived needs to be provided – were all attributes judged equally? If not need to know why and this needs to be explained – this will allow reviewers to judge the process.

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

- A downstream evaluation not done for T&E species. Rather extrapolation to other species taxonomically similar to the T&E species was provided instead. Some explanation should be provided to explain why downstream risks were not evaluated.

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 believe that the conclusions are generally supported by the analyses.

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?

- The summary discussions and conclusions were generall suitedto the original results presented. However, the final summary and conclusions will need to be revised based on the revisions recommended for specific assessment endpoints outlined above.

- The terms acceptable/unacceptable in relation to risks and risk management need to be defined in the context of the assessment, recognizing that in reality these are regulatory/risk management issue and not risk assessment issues.

- Can this ERA be used to make risk management decisions? Yes, but requires consideration of other issues, such bioaccumulation from abiotic media. In its current form, the much of available analyses are based on contaminant concentrations in biota (especially for the wildlife receptors). Remedial decisions are not generally made based on biota concentrations. As a consequence, additional modeling is needed to relate the biota-based risk conclusions to appropriate soil and sediment concentrations

- Are conclusions supported by the data – not exactly in the report as it currently is. However, after revisions in response to panel comments are completed it should.

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 data used and evaluated in this ERA is very extensive and comprehensive. I think that due to the combined efforts EPA and GE, virtually al of the relevant data and studies have been identified and incorporated. I am aware of only one study that has recently come out that should be evaluated. This is a 2003 EPA ORD report evaluating avian toxicity data for dioxin-like compounds. The full reference is:

U.S. EPA. 2003. Analyses of laboratory and field studies of reproductive toxicity in birds exposed to dioxin-like compounds for use in ecological risk assessments. U.S. EPA National Center for Environmental Assessment, Office of Research and Development. NCEA-CIN-1337. 51 pp.