SCIENTIFIC PEER REVIEW OF THE EPA REGION 2/CENAN FRAMEWORK FOR EVALUATING DREDGED MATERIAL FOR PROPOSED PLACEMENT AT THE HARS

Introduction: Goals of the peer review

The August 29, 1997 Final Rule, Simultaneous De-designation and Termination of the Mud Dump Site and Designation of the Historic Area Remediation Site, specifies that the historic area remediation site (HARS) will be remediated with uncontaminated dredged material (i.e., dredged material that meets current Category I standards and will not cause significant undesirable effects including though bioaccumulation; hereinafter referred to as "Remediation Material"). The rule further specifies that the HARS will be managed so as to reduce impacts within the Priority Remediation Area (PRA) to acceptable levels in accordance with 40 CFR 228.11. Placement of dredged material within the PRA is restricted to Remediation Material. This material will not cause significant undesirable effects, including through bioaccumulation or unacceptable toxicity in accordance with 40 CFR 227.6.

Evaluation of proposed dredged material regarding unacceptable toxicity is clearly defined in the Green Book as statistical criteria which require no interpretation. Evaluation regarding significant undesirable effects including through bioaccumulation requires assessment of chemical analyses of tissue from 28-day bioaccumulation tests. There are no specific regulatory criteria for this evaluation; however there are existing regional guideline values that have been developed and used, by the U.S. Environmental Protection Agency (EPA) Region 2 and the U.S. Army Corps of Engineers New York District, to evaluate the constituents in accordance with 227.6.

This peer review charge is to assess whether the testing evaluation process is adequate to properly determine whether a tested sediment is suitable for Remediation Material as defined. Your review should focus on the framework for evaluation of bioaccumulation data and guideline values used; it should not deal with toxicity/mortality testing. Please bear in mind that the testing evaluation applies to risks pertaining to ocean placement of the sediment, and not to risks pertaining to other alternatives such as leaving the sediment in place.

This charge is in the form of questions on critical aspects of the evaluation framework. General references are cited in each charge question to aid in finding the issue in question. Note that these are general guiding referrals and should not be considered the only review item for those specific issues. Please answer the assigned questions as directly as possible, given the provided materials and your own expertise. If you are unable to answer a particular question on the basis of the provided materials, please inform us of information needed to answer the question. Also, keep in mind that there are additional environmental data resources and test data pertaining to the New York Bight available in EPA Region 2, if they are needed.

Michael Newman, PhD.

Question 1

Is the EPA Region2/CENAN Framework for evaluating bioaccumulation results scientifically appropriate for determining the suitability of dredged material as Remediation Material? If not, describe deficiencies. (Please see Region2/CENAN joint evaluation memorandum, Figure 1)

General Answer

The framework (Figure 1 and associated text) seems generally acceptable and thoughtful. It is surprising to not see at the top of the flow diagram a statistical comparison of bioaccumulation in key endemic species inhabiting sediments of the dredge material source site versus reference sites. This seems so much more straightforward than using the 28 day bioaccumulation test and associated gross adjustments to estimate "steady state" concentrations. It also is more straightforward than comparison of the bioaccumulation test results to bioaccumulation of biota from sites near the HARS (step c, see page 14, section (ii)).

Several resolvable points limit the utility of the present framework. These points will be discussed later in this review. They include the following:

1. Step a (Chemical "X" Greater than Reference?): There are unanswered questions regarding sampling techniques (gut clearance, size/age normalization or control), specific hypothesis tests applied (page 8, paragraph 1), and ambiguity regarding test design and power estimation. These crucial aspects are generally discussed and highlighted in the Guidance Manual (Ref. 32) but specific and important details are not mentioned in the Memo of Record (Peer Review Version). Without specific details, it would be impossible to assess the true value of the information in the Memo of Record.

   Also methods associated with determining average concentration if the data set contains below detection limit observations are not specified but can greatly influence the validity of hypothesis tests. I assume that the methods described in reference 13 were used. The methods described in reference 13, "New Approach to Handling "non-detects" in Elutriate and Tissue Data" are invalid. A general statement is made that the approach was "conservative" without further explanation. The tabulated results and footnotes seem to imply that sites with all "<DL" observations or some "<DL" observations were treated differently. How were they treated? Were they treated as suggested in Ref. 13?

   Those with all "<DL" observations for the reference site and detectable amounts for the sample of interest could be analyzed by testing whether the mean for the site of interest was significantly different from the detection limit, e.g., a simple z statistic would do this (below) or simple resampling methods.

$$z = \frac{\bar{x} - DL}{\sigma/\sqrt{(n)}}$$

If some, but not all, observations were "<DL" then other tests are appropriate. Gilbert (1993, 1995) provides details on hypothesis tests for data sets with"<DL" observations including slippage, quantile, Wilcoxon rank sum, and Gehan tests. Regardless, it is important to understand that performing hypothesis tests after substituting 0, 1/2 DL or the DL for the "<DL" observations produces invalid results that are not necessarily "conservative." (Reference 13 states that 1/2 DL or 0 should be used depending on the situation. This will produce statistically misleading results. See Newman (1995) or references therein for correct methods.)

Would a high concentration (e.g., 5 times higher) be ignored if the results of the hypothesis testing for difference in sediment means had a p of 0.06 instead of 0.05 or less?

Hypothesis tests as described attempt to minimize type I error (e.g., falsely concluding that bioaccumulation was different when it was not). Hypothesis tests are commonly applied as described here but such application is inconsistent with what you are really trying to accomplish in a risk assessment. The focus should be on controlling Type II error rate and conducting tests with high power. The type II error (e.g., minimizing the probability that you will accept the hypothesis of no difference in bioaccumulation when there is a difference) and power (e.g., the ability to note a difference in bioaccumulation when there is one) would seem a more appropriate focus. For this reason, much more detail is needed regarding Type II error and power. Did you consider bioequivalence testing (see Dixon and Garrett 1993, McDonald and Erickson 1994, Dixon 1998) which handles Type II errors more appropriately?

2. The means for extrapolation from 28 day bioaccumulation concentrations to [practical] steady state concentrations are compromised. There are too many assumptions and gross approximations forced on the user.

   I assume that the 28 day bioaccumulation test is required and answers in this review should focus beyond the shortcomings of the test. However, I believe that any test which looks at a suite of contaminants simultaneously for a predetermined time will never be fully satisfactory for all contaminants. Large groups of toxicants will have inferior data as a consequence of not coming close to steady state, insufficient sample numbers, and problems of incomplete gut clearance or significant elimination during the time allowed for gut clearance. The gross manipulations to the resulting data as described here will not provide accurate information regardless of the existence of a standard test method and a history of regulatory application.

3. Means of estimating final bioaccumulation for Cd and Hg which are assumed to "violate" steady state kinetics. Inaccurate statements such as "Cadmium and mercury are not regulated in marine organisms as are essential metals, and, thus no adjustment for steady state is applicable" limit the validity of the conclusions. Regulation has little to do with coming to a steady state condition within 28 days. The final conclusion to use 28 day data for Cd and Hg as a consequence of their "non-

steady state" behavior is not logical. The concentrations increase through time with no apparent leveling off at a "steady state" concentration. They do not stay at the 28 day concentrations as a consequence of their "non-steady state" behavior.

4. No analysis of *in situ* bioaccumulation data from the proposed dredged sediment sources and reference sediment sites.

5. Use of "log K_{OW} of approximately 4 or greater" as the only cut-off for bioaccumulation [biomagnification?] (page 6, last paragraph). It isn't clear when you are discussing bioaccumulation or biomagnification. If you are discussing bioaccumulation, the cut-off point is not appropriate. If you are discussing potential for biomagnification, the cut-off is reasonable. It is consistent with Connolly and Pederson (1988) and Thomann (1989). However there can be an upper limit too as molecular size begins to influence diffusion rates. Please see Gobas et al. (1986) for more detail.

Question 2

Which of the risk-based values derived constitute "true" conservative estimates of risk levels (i.e., exceeding the value should be interpreted as sufficient cause to conclude that significant undesirable effects may result through bioaccumulation)? Which of the risk-based values derived constitute conservative screening values (i.e., test tissue concentrations below the value can confidently be interpreted to pose no risk of significant undesirable effects and exceeding should be further evaluated before the probability of significant undesirable effects can be assessed)? How can the "true" risk levels be calculated for those compounds which you believe only to have screening values? How should test concentrations be compared to risk-based levels to determine whether they are exceeded.

Answer (General answers will be given to the subquestions within this question and then specific points provided relative to each option.) Specifically, "The toxicological significance of this bioaccumulation was evaluated by: i) consideration of steady-state bioaccumulation and food-chain transfer; ii) comparison to background tissue concentrations; iii) consideration of potential ecological effects; and iv) consideration of potential carcinogenic and non-carcinogenic effects to human health" (page 12).

General Answer

First Subquestion: Which of the risk-based values derived constitute "true" conservative estimates of risk levels (i.e., exceeding the value should be interpreted as sufficient cause to conclude that significant undesirable effects may result through bioaccumulation)? In my opinion none of these methods produces a conservative estimate of risk levels that is consistent with either of the qualifiers, true or sufficient. Using the 28 day bioaccumulation test as the foundation to derive these values is problematic. <u>If you insist on the assumption that this test is valid</u>, I would suggest that iii and iv are conservative estimates of the true risk levels.

Second Subquestion: Which of the risk-based values derived constitute conservative screening values (i.e.,

test tissue concentrations below the value can confidently be interpreted to pose no risk of significant undesirable effects and exceeding should be further evaluated before the probability of significant undesirable effects can be assessed)? In my opinion none of these methods produces a "true" conservative screening value. Using the 28 day bioaccumulation test as the foundation to derive these values is problematic. If you assume that this test is valid, I would suggest that i to iv combined provide a conservative estimate of the true risk levels.

Third Subquestion: How can the "true" risk levels be calculated for those compounds which you believe only to have screening values? In my opinion, I would use the concentrations in representative species living at the site from which the dredged materials are to be taken and use them in the comparisons described, i.e., use them in approaches (i) to (iv). If there were relevant commercial/sports species or major forage species for commercial/sports species relevant to the source and HARS sites, I would consider them. I would consider biomagnification when reasonable. For example, lobster stocks at the HARS are noted as having high TCDD/PCB levels. Lobsters are long lived and could easily biomagnify contaminants such mercury from benthic forage species and act as a vector to human exposure. After such consideration, I would then apply (ii) to (iv) with possible modifications as described below.

Fourth Subquestion: How should test concentrations be compared to risk-based levels to determine whether they are exceeded? Under the assumption that the 28 day bioaccumulation test is valid, use the method described in the answer to "Third Subquestion" including the modifications described below, e.g., additional consideration of a hazard index.

Specific Points

(i) Consideration of Steady-state Bioaccumulation and Food-Chain Transfer

The definition of steady-state as "the lack of significant difference (ANOVA, alpha=0.05) among tissue residues taken at three consecutive sampling intervals" is not ideal. It is particularly problematic if not linked to a required type II error rate, level of power or minimum significant difference. Statistical power must be considered in order to get a meaningful decision from an ANOVA. For example, wide variability within times could result in the acceptance of the null hypothesis of no difference in means when there was a difference. Sampling three very close time intervals could also result in an acceptance when there was a difference.

Metals: The statement that metals bioaccumulate faster than organic compounds is not valid. For example, cadmium will slowly bioaccumulate during the life of an organism, and some organic compounds can be taken up very rapidly.

The assumption that metals do not biomagnify is reasonable with the major exception of mercury which is a classic example of a toxicant that does biomagnify. The use of a trophic transfer coefficient of one is not appropriate for mercury. See Cabana and coworkers (Cabana and Rasmussen 1994, Cabana et al. 1994) as an example. See also third paragraph on page 11 of reference 57.

Pesticides and Industrial Chemicals and PAHs: General extrapolations from 28 day bioaccumulation levels to steady state does not seem appropriate.

(ii) Comparison of test results to background tissue concentrations

Here concentrations in organisms taken from near (but not in) the HARS are compared to those from the bioaccumulation tests. The comparison of these organisms to those from a 28 day test that may not be at steady state seems dubious. Why not compare animals from near the HARS to those endemic to the site from which the dredged materials are to be taken?

(iii) Consideration of potential ecological effects

A literature review was done to evaluate the bioaccumulation test results relative to ecological effects. The CV is multiplied by an estimated BCF and compared to that concentration "protecting" 95% of tested species in the database. Except for the paucity of data from which to draw conclusions and the dubious extrapolation from the bioaccumulation test, this is a reasonable indicator. Critical body residues were used for PAHS which also seems plausible.

(iv) Consideration of Potential Carcinogenic and Non-Carcinogenic Effects on Human Health

The focus here is risk to humans by ingestion. As discussed above, the extrapolated "steady state" concentrations are suspect and, therefore, the "human cancer protection levels" derived from them are also suspect.

The hazard quotient was used for non-carcinogenic effects. Would it be reasonable and informative to also extend this approach to the calculation of a hazard index (sum the hazard quotients) under the assumption of additivity? Biomagnification to a commercial species foraging on these representative benthic species is inappropriately ignored in the analysis, e.g., mercury biomagnification.

Question 3

In conducting the integrated effects evaluation using the types of data provided by the applicant, which of the eight factors for LPC compliance listed in the Green Book are appropriate and relevant? How can a quantitative/strategic framework be established to evaluate tissue data for those factors? Considering that comparison to regional Matrix values and site-specific risk values represent case-specific evaluations, is it necessary to conduct the integrated effects evaluation of the bioaccumulation results? (Please see Reference No. 61, page 6-6)

Subanswer One: In conducting the integrated effects evaluation using the types of data provided by the applicant, which of the eight factors for LPC compliance listed in the Green Book are appropriate and relevant? These eight factors being:

1. Number of species in which bioaccumulation from the dredged material is statistically greater

than bioaccumulation from the reference material,
2. Number of contaminants for which bioaccumulation from the dredged material is statistically greater than bioaccumulation from the reference material,
3. Magnitude by which bioaccumulation from the dredged material exceeds bioaccumulation from the reference material,
4. Toxicological importance of the contaminants whose bioaccumulation from the dredged material statistically exceeds that from the reference material,
5. Phylogenetic diversity of the species in which bioaccumulation from the dredged material statistically exceeds bioaccumulation from the reference material,
6. Propensity for the contaminants with statistically significant bioaccumulation to biomagnify within aquatic food chains,
7. Magnitude of toxicity and number and phylogenetic diversity of species exhibiting greater mortality in the dredged material than in the reference material,
8. Magnitude by which contaminants whose bioaccumulation from the dredged material exceeds that from the reference material also exceed the concentrations found in comparable species living in the vicinity of the proposed disposal site.

It is my opinion that they are all relevant in the context they are presented. Their relative importance can be crudely ranked: 1=2=3=4=6>8>7>5.

Subanswer Two: How can a quantitative/strategic framework be established to evaluate tissue data for those factors? The impression given in the Green Book is that these items are to be used in the context of expert opinion, not a structured decision flow chart. I would use review by experts, perhaps based on the crude rankings of priority provided above, as a means of incorporating these aspects of the assessment.

Subanswer Three: Considering that comparison to regional Matrix values and site-specific risk values represent case-specific evaluations, is it necessary to conduct the integrated effects evaluation of the bioaccumulation results? Yes. I believe that the integrated effects evaluation provides applicable information, although it would be much more useful if it did not rely so heavily on the 28 day bioaccumulation test and associated extrapolations. According to the information provided (e.g., ref. 57), regional matrix values exist for only a subset of contaminants. (Please note that I did not find ref. 61 in my review package so I could not respond specifically using information from page 6.6 as suggested.)

Benchmark and Risk Evaluation Values

Regional Matrix Values

Question 4

Are the Matrix values suitable for determining the suitability for placement at the HARS as Remediation Material?

Answer based on Ref. 57 and the Peer Review Version. Review of Compliance with the Testing Requirements of ...: Yes. The values seem appropriate as gross estimates if used as originally planned. "This guidance is considered by the Corps to be dynamic. It is not established as fixed and final numerical criteria but rather as an aid to interpreting test results. As such, it is open to review and updating as additional pertinent data become available" (Ref. 57, 1981).

Question 5

Regional Matrix values were developed in 1981 by compiling available field data for mercury, cadmium, PCBs, and total DDTs. Were these values derived appropriately for their intended use? Based on current data sets and scientific literature, are these 1981 values suitable for predicting the significant undesirable effect due to bioaccumulation? (Please see Reference No. 57) If not, identify more current references, data sets, and/or actual chemical specific values that would be more appropriate.

Answer: The information gives general estimates as intended. As noted above, extensive review and augmentation of this information should be done. Since the evaluation was last conducted in 1981, a new review is due.

Regional Dioxin Values

Question 6

Currently, the presence of 2,3,7,8-TCDD at a detectable concentration (i.e., greater than or equal to one part per trillion (pptr)) in tissues of organisms exposed to dredged material precludes its classification as Category I (hence Remediation Material); presence of the remaining dioxin/furan congeners, at concentrations of TEQs equal to or greater than 4.5 pptr, results in a similar conclusion. When 28-day tissue concentrations exceed these values, is there sufficient cause to conclude that placement of the material is not suitable as HARS Remediation Material? If not, what levels indicate sufficient cause for this conclusion? (Please see Reference No. 89)

Answer: I would tend to agree with the present decision process.

Question 7

Are dioxin values suitable for predicting the significant undesirable effects due to bioaccumulation? If not, should these values be based on a risk analysis paradigm in which the size of the human population subgroup potentially exposed through intentional behavior is compared to the size of the general population in the EPA? Since the primary route of exposure is through consumption of fish and shellfish, should the variability in potential exposure due to differences in fishing behavior (e.g., target species, seasonal preferences) be incorporated in the risk paradigm? How would a benchmark protective of human health compare to benchmarks determined using an ecological risk analysis paradigm for resident fish and piscivorous wildlife?

Subanswer One: Are dioxin values suitable for predicting the significant undesirable effects due to bioaccumulation? Yes. They seem suitable.

Subanswer Two: Since the primary route of exposure is through consumption of fish and shellfish, should the variability in potential exposure due to differences in fishing behavior (e.g., target species, seasonal preferences) be incorporated in the risk paradigm? Yes. I think that at a certain tier in the risk assessment, more detailed information would be useful. The present analysis seems to focus on the general population but, perhaps, should also consider subpopulations of humans with higher consumption rates of fish and shellfish. Inclusion of fishing behavior as mentioned above could then be added to the analysis.

Subanswer Three: How would a benchmark protective of human health compare to benchmarks determined using an ecological risk analysis paradigm for resident fish and piscivorous wildlife? The benchmarks for humans should be focused on protecting individuals but those for fish and wildlife should be based on maintaining viable populations. Therefore, the general response would be that the benchmarks for humans should be lower than those for fish and wildlife assuming that one adjusts for differences in exposure characteristics.

FDA Action Levels (Please see Reference No.61, Sec. 6.3)

Question 8

Are FDA Action Levels useful as upper limit human health benchmarks?

Answer: Yes. The FDA Action Levels are useful.

Question 9

Would the evaluation be improved by omitting comparison of tissue results to FDA Action Levels?

Answer: No.

7. Human Health Risk, Cancer and Noncancer

Question 10

A. Are the risk values suitable for determining the suitability for placement at the HARS as Remediation Material? If there are better alternatives for human risk, specifically what are they?

Answer: The process seems reasonable.

Question 11

B. Benthic tissue levels for cancer protection were derived using assumptions focused on attaining a cancer protection at the 10-4 risk level. Is this risk appropriate for a determination of ocean placement of Remediation Material? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-4, A-5)

Answer: The obvious range of candidate risk levels is 10^{-4} to 10^{-6}. I would tend to agree with the 10^{-4}

used here. However, I am biased towards doing these calculations for subpopulations (e.g., groups prone to consume more fish/shellfish from the area), not the general population.

Question 12

C. Benthic tissue levels for noncancer protection were derived using Reference Dose (RfD) of several organic and inorganic contaminants for the protection of human health. Are these values appropriately and consistently derived? Is the whole body/fillet conversion factor of 1.35 an appropriate factor for all of the contaminants considered if human exposure is assumed to be primarily via consumption of the fillet portion of the fish? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachments B and C) If not, what factors would be appropriate? For the lead noncancer value, since there is no RFD for lead the value was derived differently than the other metals. Was the value derived appropriately? (Please see Reference No. 88)

Subanswer One:Are these values appropriately and consistently derived? I see no general problem.

Subanswer Two: Is the whole body/fillet conversion factor of 1.35 an appropriate factor for all of the contaminants considered if human exposure is assumed to be primarily via consumption of the fillet portion of the fish? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachments B and C) If not, what factors would be appropriate? This seems appropriate.

Subanswer Three: For the lead noncancer value, since there is no RFD for lead the value was derived differently than the other metals. Was the value derived appropriately? (Please see Reference No. 88). The use of data for blood lead levels in urban children seems conservative (and therefore appropriate). They often tend to be very close to background levels. The calculations used high drinking water (4 ppb) and dust/soil/paint (800 ppm) lead levels, giving the entire process a conservative bias. It is unlikely that fish will be consumed by a child as specified in ref. 88. Again this results have an appropriately conservative bias. The final calculated limit of 1.3 ppm in fish tissue sounds reasonable.

Question 13

Are the risk values suitable for predicting the significant undesirable effects due to bioaccumulation? Since the primary route of exposure is through consumption of fish and shellfish, should the variability in potential exposure due to differences in fishing behavior (e.g., target species, seasonal preferences) be incorporated in the risk paradigm?

Answer: This is a good point. As suggested elsewhere (Answers to Questions 7, 11 and 26), it would be helpful to examine subpopulations at higher tiers of the assessment process. Are there appropriate data available from surveys of fishing and fish consumption practices for the area?

8. Ecological Risk

Question 14

Ecological effects benchmarks include the Water Quality Criteria Tissue Level (WQCTL), Critical Body Residue (CBR) associated with narcotic responses, and certain mutagenic/teratogenic effects. Is it valid to use the CBR effect end point for evaluating significant undesirable effect? Are there other ecological end points that should be used to measure ecological risk that are protective of marine benthic and fish life via trophic transfer, particularly for PAHs? If so, identify. With regard to a narcotic effect for chlorinated organic compounds, should an additive approach be considered to include the contribution of chlorinated hydrocarbons against this narcotic (CBR) endpoint.

Subanswer One: Is it valid to use the CBR effect end point for evaluating significant undesirable effect? Are there other ecological end points that should be used to measure ecological risk that are protective of marine benthic and fish life via trophic transfer, particularly for PAHs? If so, identify

If one assumes that the 28 day bioaccumulation test results are valid, then the use of CBR is valid. There are some toxic effects that will not conform to the CBR concept, e.g., cumulative liver damage by cyanide. However, the application of the CBR here seems the best way to estimate the potential for effect. Relative to ecological risk, there are many other end points that could be used. One could survey the site from which the dredge material will come and measure many things. One could measure community indices, lesion incidence, demographic qualities of key species, biochemical qualities of individuals, or a variety of other conventional metrics.

Subanswer Two: With regard to a narcotic effect for chlorinated organic compounds, should an additive approach be considered to include the contribution of chlorinated hydrocarbons against this narcotic (CBR) endpoint. Yes. An additive effect seems logical and consistent with this mode of action.

Question 15

Is the EPA 2 WQCTL approach (i.e., multiplying the Water Quality Criteria Chronic Value by the Bioconcentration factor) appropriate for determining ecological effects levels of the contaminants for which they were developed? Specifically, are the appropriate BCFs used (for fish, bivalves, etc)? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-1)

Answer: The application of a single BCF from a representative organism is an appropriate gross metric. But there is no single BCF for a species or a single representative species. Likely what was used was the closest species for which there was a published BCF. Consequently, I would apply an uncertainty factor (UF=10?) in this calculation.

Question 16

BCFs reported for fish were used in the calculations of WQCTLs for organics; is this derived level appropriate for setting benthic tissue ecological effects levels? If the fish tissue levels are used, should adjustments be made to the derived levels to reflect the higher lipid

contents of the benthic organisms used in the testing program?

Answer: No. A BCF for fish is not appropriate to use for an invertebrate. Further the BCF for one group of invertebrates may not be appropriate for another invertebrate group. Too many factors have large effects on BCF to make such a general use of a single BCF. This seems to be a situation in which an uncertainty factor is appropriate.

Question 17

Are the WQCTLs calculated for metals using bivalve BCFs appropriate for setting levels for polychaetes or vice versa?

Answer: No. Please see the answer above.

Question 18

Are the uncertainty factors applied while deriving ecological effects levels for PAH contaminants appropriate? Does this adequately address the uncertainty around the derived values? Can uncertainty be accounted for using these order of magnitude adjustments? Should they be applied elsewhere to the other risk-based values?

Answer: Relative to converting effect to no effect, the associated calculations contain considerable error. I would suggest an uncertainty factor of 10. The means of dealing with narcotic effects seems adequate. An interspecies uncertainty factor of 10 seems adequate.

Question 19

Are the risk values suitable for predicting the significant undesirable effects due to bioaccumulation; are there better alternatives for ecological nonspecific risk?

Answer: There are additional and more direct means of assessing ecological risk. Conventional metrics of effects to individuals (e.g., lesions), populations (e.g., demographic assessment or abundance estimation) and communities (e.g., species richness, IBI, and others) could be done at the site from which the material is to be taken.

Question 20

If you believe that these values are over- or under- conservative, what do you believe to be an appropriate way to improve them?

Answer: I would guess that they are overly conservative. However, there is so much uncertainty in several steps that it is difficult to be more definitive. Please see the above comments for possible improvements. It is unfortunate that the approaches described do not provide any direct information on population viability or benthic community effects.

Calculations

Question 21

Should total PCBs continue to be estimated by doubling the total of 22 congeners or should it be quantified directly using another measure of quantification? What method is most appropriate for sediments in the NY/NJ Harbor area? (Please see Reference No. 60, Table 4-4B)

Answer: I am not qualified to answer this question.

Question 22

Currently, 28-day tissue concentrations of certain organic contaminants are adjusted by some multiplier to estimate the concentrations of those compounds had the exposure been of sufficient duration to allow attainment of steady state levels. (Please see Reference Nos.5 and 46) Are these adjustments appropriate? Should steady state corrections be applied to any other of the listed contaminants? Are there other compounds for which we test that are not expected to approach steady state within the 28-day period?

Answer: This is my major concern with the process. I believe that such use of multipliers will not provide an accurate measure of steady state concentrations for many of the toxicants of concern. Therefore all decisions based on these numbers are compromised. By using one test, one will not be able to fit simple models for accumulation through time and use the model parameters to estimate practical steady state concentrations of all toxicants of concern. Optimal design for some contaminants relative to the spacing of sampling through time and number of individuals required at each sampling would render such a test compromised or inadequate for other contaminants. This is probably the reason for defaulting to multiplying the 28 day bioaccumulation test concentration by some factor. Regardless, I believe that the multipliers are not the solution to getting steady state concentrations. The cited references provide conditional information that may not hold in other situations (e.g., sediments with very different organic carbon or AVS content, species with different lipid contents, species with different detoxification/sequestration mechanisms, or different temperatures or feeding rates). The cited references deal with organic compounds only. Is there sound evidence to support the universal utility of these multipliers? Is there evidence that they are grossly universal in their utility? Were these values to be checked against the realized concentrations in benthic species at the site from which the dredge material will be taken (as suggested in ref. 5)?

Some results from various publications are applied incorrectly, i.e., using 28 day bioaccumulation for Hg and Cd because they do not conform to the steady state context. Wouldn't it be more accurate to estimate the concentration that would be reached during the average life span of the organism since the Cd or Hg concentrations continue to rise with time and do not appear to come to a steady state concentration?

Question 23

Is the calculation and use of BaP toxicity equivalence an appropriate way to estimate the potential carcinogenicity of PAHs? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Section C.)

Answer: The calculations and use of toxic equivalence appear adequate.

Question 24

Similar to PCBs, only a subset of those PAHS present in New York Harbor are measured for testing evaluation. How should the remainder be considered?

Answer: I lack sufficient background to give an informed answer.

Question 25

Is the assumption of a trophic transfer coefficient of one appropriate for use in evaluating the potential for human health and ecological impacts associated with metals in Remediation Material? Are the trophic transfer factors calculated for organic compounds correct? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachment C.)

Answer: The trophic transfer coefficient of 1 seems conservative (i.e., it could be less than 1 for some metals) except for mercury. Because mercury will biomagnify, the coefficient greater than this is required. Please see the references cited below for the appropriate coefficient for mercury.

Question 26

Is the assumption of a fish consumption rate of 6.5 g/day appropriate for use in evaluating the potential for human health impacts associated with metals in Remediation Material? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-5) Would it be appropriate that the evaluation focus on a higher consumption population?

Answer: The consumption rate is consistent and reasonable if the focus is the general population. However, as suggested in EPA 540/1-89/002 December 1989, "Residents near major commercial or recreational fisheries or shell fisheries are likely to ingest larger quantities of locally caught fish and shellfish than inland residents." I would suggest that consumption surveys be reviewed to identify any possible subpopulation consuming higher amounts than 6.5 g/day.

General

Question 27

Is it plausible to replace any other risk assessment assumptions with assumptions specific to the HARS site? (Please see Region2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachment C and Reference Nos. 88) Is it appropriate to consider the HARS intended use to be factored into an evaluation of effects at the community or population level?

Answer: I believe that effects to populations and communities are appropriate. Please see those points already discussed relative to community and population level effects.

Question 28

Is use of the Squibb et al. (1991) report appropriate for identifying the contaminants of concern? Are there contaminants which should be added to or deleted from the list of contaminants for which we presently test? Please see Reference No. 51)

Answer: The report seems appropriate. Perhaps organotins such as TBT should be added.

Question 29

Should risks from synergistic effects, from exposure to multiple contaminants, be evaluated using results from tissue analyses? If so, how? If not, why not?

Answer: This is a good point that has no clear answer. Synergism should be considered if there are high concentrations of specific toxicants known to have synergistic effects. However, additivity is likely the best option if such information is lacking.

Question 30

Is test tissue concentration exceeding reference tissue concentration by less than 10X a meaningful evaluative criterion? (Please see page 9 of the Region2/CENAN joint evaluation memorandum)?

Answer: No. If pressed to apply such a rule, I would express exceedance in the context of standard deviations above the mean of the reference. (Also please remember that some of the difficulties with applying these types of comparisons is the invalid use of "<DL" observations.) The more important point is whether the tissue concentrations are close to or above some toxicological threshold.

Question 31

Are the studies from which background tissue concentrations were calculated weighted appropriately? If not, what method is recommended? Is the use of the mean the most appropriate measurement of central tendency? If not, what measure should be used? (Please see Reference No. 98) Are the assumption, presented on page 14 pertaining to comparisons of bioaccumulation in test tissue to tissue concentrations in organisms from the vicinity of the remediation site, valid for evaluating undesirable effects?

Answer: The weighting was adequate. Again, the treatment of "<DL" observations is invalid, i.e. substitution with 1/2DL (see Newman (1995) for details). Relative to the use of the mean, distributions of contaminant concentrations generally conform to a log normal distribution more often than they conform to a normal distribution. Therefore, the arithmetic mean is probably not the best estimate of central tendency. I would test for differences using log transformed concentrations, not arithmetic concentrations. Relative to comparisons of background tissue concentrations to those estimated in the 28 day bioaccumulation test, I feel that such a comparison has limited value. I would prefer to compare the concentrations in biota at the material source to concentrations in biota near the HARS.

Question 32

Can baseline tissue concentrations, from appropriate benthic organisms resident to the HARS, be used as standards to determine suitability for Remediation Material as defined above?

Answer: Yes. This has been mentioned in several of my answers above.

References

Cabana, G. and J.B. Rasmussen. 1994. Modelling food chain structure and contaminant bioaccumulation using stable nitrogen isotopes. Nature (LONDON) 372: 255-257.

Cabana, G., A. Tremblay, J. Kalff and J.B. Ramussen. 1994. Pelagic food chain structure in Ontario lakes: A determinant of mercury in lake trout (*Salvelinus namaycush*). Can. J. Fish. Aquat. Sci. 51: 381-389.

Connolly, C.D. and C.J. Pedersen. 1988. A thermodynamic-based evaluation of organic chemical accumulation in aquatic organisms. Environ. Sci. Technol. 22: 99-103.

Dixon, P.M. 1998. Assessing effect and no effect with equivalence tests. in *Risk Assessment. Logic and Measurement*, M.C. Newman and C.L. Strojan (Eds.), Ann Arbor Press, Chelsea, MI.

Dixon, P.M. and K.A. Garrett. 1993. Statistical issues for field experimenters. in *Wildlife Toxicology and Population Modeling Integrated Studies of Agroecosystems*, R.J. Kendall and T. Lacher (Eds.), Lewis Publishers, Chelsea, MI.

Gilbert, R.O. 1993. Letter of July 30, 1993 to Beverly Ramsey, Systematic Management Services, Inc., Germantown, MD.

Gilbert, R.O. 1995. A review of statistical methods for data sets with multiple censoring points, Report PNL-10652. Battelle Pacific Northwest Laboratories, Richland, WA.

Gobas, F.A., P.C. Opperhuizen and O. Hutzinger. 1986. Bioconcentration of hydrophobic chemicals in fish: Relation with membrane permeation. Environ. Toxicol. Chem. 5: 637-646.

McDonald, L.L. and W.P.Erickson. 1994. Testing for bioequivalence in field studies: Has a disturbed site been adequately reclaimed? in *Statistics in Ecology and Environmental Monitoring*, D. Fletcher and B.F.M. Manly (Eds.), Univ. of Ontago Press, Dunedin.

Newman, M.C. 1995. *Quantitative Methods in Aquatic Ecotoxicology*. Lewis Publishers, Boca Raton, FL.

Thomann, R.V. 1989. Bioaccumulation model of organic chemical distribution in aquatic food chains. Environ. Sci. Technol. 23: 699-707.