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

Peer Review Comments
John H. Gentile, PhD
University of Miami
August 24, 1995

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 that 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 on 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.

Framework

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)

Answer: Fundamentally the framework outline in the CENAN joint evaluation memo and illustrated in Figure 1 contains the necessary elements to make an evaluation of the potential health and ecological effects of contaminant bioaccumulation. The only apparent deficiency is not with the framework *per se* but with the lack of estimates of variability that support the statistical analysis for determining when there is significance exceedence of a reference value or the various benchmarks that are used. I am not familiar with how the many assumptions that are implicit in the Framework were arrived at but I will assume that they all have been peer-review for their statistical soundness. I do like the additional 8-Factors but would have to see a case where they played a dominant role before I would comment on the limits of their applicability.

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: Currently, you've identified the 'true' conservative estimates of risk to be: FDA Action Levels; Matrix level; or Dioxin Category 1 value from the perspective that exceeding these values classifies the dredged material as not being Category 1.

A conservative screening value could be set at no significant difference from reference as long it satisfies several assumptions:1) the reference values are representative of uncontaminated sites throughout the country and not just locally; 2) that the reference values have been shown to be associated with healthy benthic community structure and function or from laboratory studies shown to have no associated biological effects; and 3) that the statistical design for determining significant differences satisfies assumptions of random sampling, proper selection of replicates and avoids the issue of pseudo-replication.

The comparison of risk-based level to test concentrations gets at the heart of my concerns with the application of this framework. That is, is there sufficient replication in the determination of test values, reference values, WQCTLs, FDA and Matrix benchmarks to determine statistical differences with prescribed levels of confidence? If not then this framework is not risk-based and should not be purported to be.

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)

Answer: It is interesting that you are asking the peer-reviewers this question when that was exactly what I referred to in my comment on the suitability of the Framework. I was anticipating that you would provide an example of the application and then ask our evaluation rather than the other way around. Nevertheless, there are several ways to stratify the 8-Factors into a decision framework based on some set of criteria as long as it is recognized that these are inter-related to varying degrees. One approach is to separate the eight into 1st and 2nd order factors. Personally if I were a manager, I'd like the 1st-order factors to tell me if I have a potential problem. These could include: the toxicological importance of the contaminants; the magnitude of the bioaccumulation in any one or more species; and the propensity to biomagnify in food webs. The latter is particularly important since most of the effects we see are detected in top predators be they mammals, birds, or reptiles.

Given I have a potential problem I'd want to know how serious it was. The 2nd-order factors provide supporting evidence to reduce potential uncertainty and further define the magnitude and extent of the problem. They could include the number of species; number of contaminants; phylogenetic diversity; the magnitude to values for species in the disposal site.

Personally, I do not think the second question is an appropriate one for a peer-review, that is develop a quantitative strategic framework using these eight factors. That was and should be the responsibility of EPA/COE and its contractors. Having said that my colleagues and I have developed an ecological significance decision framework using several factors that might be relative though we did not attempt to quantify it that is due to be published in the August issue of Human and Ecological Risk Assessment (HERA). I'd be happy to provide a reprint when available.

I do think there are situations where there is a need to conduct the Integrated Effects Evaluation (IEE). For example, the Regional Matrix only accounts for two metals and two organics - what about the other contaminants. Second while the Risk Evaluation using WQCTLs expands the contaminants it really is based on national and not site-specific conditions. Personally, I like the IEE because it would provide considerably more information than the "point estimates" that would reduce the uncertainty associated with decisions. I'd suggest the next step is to develop a strategic plan for constructing a quantitative decision-analysis framework for using the 8-factors or more or less as need be.

Benchmark and Risk Evaluation Values

4. Regional Matrix Values

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

Answer: The fundamental problem with all point estimates or benchmarks is that they have no estimates of variability around them so there is no measure of uncertainty related to a decision derived from their application. Since a fundamental element of risk-based analysis and risk-management is decision-making in the face of uncertainty, then measuring probabilities and uncertainties would seem not only logical but also necessary. From a managers perspective I'd want to know whether the 2x, 3x or 10x exceedence was real or within the both the biological and analytical variability. I'd certainly only use it as a screening tool and not a decision tool. However, if the policy has and continues to be based on point estimates (e.g., criteria, benchmarks, etc.) then the Matrix values are sufficient.

B. 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: I was peripherally involved with this process in the 1980's while with EPA and directing a Field Verification Program with COE. At that time the derivation the concept and its limitations were debated and the consensus was that given the state of the science this would be a useful tool for screening the potential ecological effects of contaminant tissue residues in biota. I don't think at that time, we considered this a predictive tool nor do I now for that matter. The idea was that if you exceed this value then it would trigger the need for further investigation. Among the issues were whether one should DDT and PCBs in toto or look at their individual congeners. The latter was not deemed possible at that time for a several reasons; lack of data on specific congeners, reliability of analytical methods for congeners, and a lack of congener specific effects data or even potency data at that time.

I would assume that in the intervening time (1981-1998) that three things would have happened: 1) the original data base has been expanded as new data became available thus providing a much sounder scientific basis for decisions including data on PCB congeners; 2) the concept would have been expanded to include additional contaminants, particularly the PAHs; and, 3) the efficacy of the original Matrix would have been evaluated as a management tool. Not having been involved with this area of research since the mid-eighties it is difficult for me to judge but that is certainly at the heart of your last questions. That you are suggesting Reference #57 as supporting information suggests that little further research has gone into refining and re-defining the Regional Matrix. If the plan is to continue using this approach, I would encourage the Matrix concept be more risk-based, that is, utilize distributions, probabilities, and uncertainties in the decision-making framework. Further, Reference #57 highlights the problem of using a "dated" strategy when it concludes that there is no evidence that methylmercury is not a threat to bioaccumulate and biomagnify

5. Regional Dioxin Values

1. 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 their 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: Reference No. 89 provides a good summary of the policy and approach for evaluating dioxin risks Ð though it is not a risk assessment *per se* for all the reasons I've discussed above. Using point estimate benchmarks or criteria results in a hazard index type of framework and not a probabilistic-based risk framework. Nevertheless the review was informative and basically sets up an hazard index with three benchmarks; ²1, 1-10, and >10pptr. Two concerns with this approach one is with the philosophy supporting the policy and the second with implementation. First, is that TEQ value of 4.51 is based on the sum of 1/2 the detection limits for the non-2,3,7,8 Ð substituted dioxin/furans times the individual TEQs. Making policy decisions at the detection limits is problematic at best. Second, how does one decide if exceeding the 1 pptr in a 28-day test is cause for concern given that your decision framework has no way to treat variability and uncertainty. What if the a single 28-day test results in a value of 1.5 or 1.9 or even 2.6 how does one decide if this is really a problem or it is within the variability of the bioaccumulation testing and analysis methods itself particularly when as stated the values are at the detection levels. What needs to be determined is the amount of variability around the benchmark that is acceptable based upon the consequences to human health. To answer that question I'd determine how much exceedence is statistically significant and what are the consequences of that magnitude of exceedence, that is, what is the incremental health and environmental risk. If the exceedence is statistically significant and above, 3.6 for example, then the health consequences will equal or exceed accepted risk criteria deeming the material unacceptable. Some analyses like this would appear to be necessary to answer the suitability question.

B. 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?

Answer: Reference 89 addresses the first question in some detail, however, I am not convinced that the current method has real power for predicting undesirable effects but rather provides useful tool for establishing policy boundaries. A risk-based approach would be much more realistic particularly if based upon site-specific information such as at risk human sub-populations, different dietary intakes, fishing behavior, etc. Comparison of health and wildlife benchmarks would have to wait until the specific analyses were done. However, my suspicion is that the use of the risk paradigm and site-specific wildlife information

would produce a more defensible and robust risk assessment that would have less uncertainty than the health assessment.

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

Are FDA Action Levels useful as upper limit human health benchmarks? Would the evaluation be improved by omitting comparison of tissue results to FDA Action Levels?

Answer: The FDA Action Levels are of limited value since there derivation is complicated by the addition of factors such as economics and thus are not directly related to a health effect endpoint. That is, they are not effects specific, that is, coupled to teratogenic effects, mutagenic effects, reproductive effects, etc. Thus these values are not a one-to-one equivalent of effects and as such are at best a poor upper bound estimate with not estimate of uncertainty. I recall calculating the incremental risk for PCBs, and Dioxins at the FDA levels and if my memory serves me they were in the 10^{-2} to 10^{0} range which is well above the 10^{-4} benchmark. Thus you could be in compliance with the FDA Action Level and still result in an unacceptable incremental risk for cancer or some other endpoint. To me they give a false sense of security There are other benchmarks out there for evaluating the human health effects of contaminant residues that have been developed by WHO and other countries that might be more useful. Another final reason for not including them is that they are often misinterpreted as being protective of the environment.

7. Human Health Risk, Cancer and Noncancer

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: Human health risks are not my area of expertise so I can only offer general suggestions to this question. The approach and methods employed are those that are currently accepted by the regulatory and scientific communities with the exception that there are advocates for using distributions rather than point estimates and conducting Monte Carlo Simulations resulting in a distribution of risk probabilities. The controversy with this approach centers on being able to define the appropriate distribution parameters, nevertheless it is something that you need to consider as part of this analysis. As the CENAN joint evaluation memo (pages A-4 and A-5) states, this approach must be considered a conservative upper bound estimate. What bothers me about this approach from a risk perspective is that it is has little or no basis in reality for a host of reasons. For example, how does one address the issue of calculating the proportion of contaminant coming from fisheries in the HARS vs. the total catch into which the HARS sub-population will be diluted and the subsequent probability of any person in the NY/NJ region of consuming enough fish to even remotely approach the upper bound. One could make those types of estimates and create a response surface that might be very informative.

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: Based upon the rationale presented in the Appendix for Table 1, Pages A-4, A-5, I would agree that if one accepts all the assumptions and wishes to take a very conservative approach to avoid dealing with uncertainties then this is appropriate. It might be useful to provide a estimate of the probability of exceeding the cancer protection benchmark by preparing a matrix of tissue concentrations and daily intakes and their incremental risks. Then compare those values to the range of tissue concentrations from contaminated sites just to see if it plausible to experience a set of conditions that would lead to exceeding the upper bound.

C. Benthic tissue levels for non-cancer 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 non-cancer 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)

Answer: This is not my area of expertise.

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: The answer to this question is 'yes' if a conservative upper bound is the management goal and comfort level and 'no' if one wishes to insert a truly risk-based sense of reality to the problem. I touched on this in my comments above under 7A. If the goal is to develop a truly risk-based estimate of human risk the crucial information is that relative to exposure that is, what is the probability and proportion of contaminated fish or shellfish coming from the HARS site that ends up in the diet of one or more sub-populations with a range of dietary intakes. Target species and seasonal preferences are but some of the variables that need to be included.

8. Ecological Risk

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 their 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.

Answer: My familiarity with the CBR literature is limited to what I have read in the supporting materials so my comments to these questions may have limited value. Nevertheless, after reading McCarty's 1992 paper in Environmental. Toxicology and Chemistry, a few thoughts are worth noting. First the data base used to develop the CBR is derived solely from freshwater for chemicals, primarily the fathead minnow, and with chemical with log $K_{ow} > 1.5$. For this data base and a very limited number of endpoints the relationship between CBR and lethality relationship for narcotics can be approximated by the QSAR derived equation CBR(mM) = 2.4 mM +50/K_{ow}. The fathead minnow specific CBR is 4.4 (mM) with a range of 2.2-2.8 mM. Multiplying by a factor of 0.25-0.1 can approximate conversion to chronic toxicity. There are several relevant questions that need to be addressed before one can confidently apply this approach to the marine fish and invertebrates: 1) has this relationship been corroborated for marine fish and invertebrates; 2) has the CBR approach been widely applied after the original work of McCarty; 3) have alternative hypotheses for the PAHs been proposed (e.g., Swartz et al. 1996); and 4) has any confirmatory studies been done to further develop the chronic relationship? What I noted in the CENAN memo was that there were no alternatives presented to the CBR and no literature cited beyond McCarty's original work. This tells me that either there is ' no better show in town' or no one has looked at the recent literature. I am not familiar enough with this area but I'm sure other peer reviews will provide useful information. My final comment is that if a sound argument can be made for transference of this approach to marine species based upon first principles and mechanisms then one could use it as an index much like the other benchmarks that you've chosen. However, until all the assumptions have been tested then I'd certainly be very cautious - if you don't need it then don't include it.

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: Yes, given there are no studies that explicitly describe the residue-effects relationship. This approach relies on the accuracy of the contaminant-specific BCFs which based on the material submitted for review is fine.

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: As stated in the Appendix for Table 1, Section A this approach is accepted to be conservative and therefore should be protective of 95% of all tested organisms which do include not only fish but representatives from several phyla including benthic species. This approach is generic and can be

made site specific by modifying specific factors if it is deemed appropriate. The calculation can be done with and without the site-specific data to determine if the difference results in a significant change in interpretation.

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

Answer: The BCFs would have to be compared for a representative set of organic and inorganic compounds to make a judgement. There is considerably more data on metals uptake with bivalves than polychaetes so I'd use the bivalves. However, I'd try to determine if the bivlaves were consistently protective of polychaetes and when there exceptions.

Are the uncertainty factors applied while deriving ecological effect 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: The discussion of potential ecological impacts for PAHs and their uncertainties are discussed in Appendix to Table 1 (A-2 and A-3). The explanation of the derivation and variability is sufficient though a more complete discussion is in McCarty 1991 and McCarty et al. 1992. However, the choice of 40 ppm (40,000 ppb) as the value in Table 1 is based on fish and is being compared to invertebrates (polychaete and bivalve) which have much higher effect thresholds. I assume this is in keeping with a conservative approach. Regarding the uncertainty around the derived values, McCarty et al. 1992 reports that the range of concentrations causing narcotic effects on aquatic organisms is from 1.4 to 21*u*moles/g wet weight which is a factor of 15. Thus an appropriate 'safety' factor to account for differences in species-species sensitivity and to protect for untested species could be set at 10-20 for freshwater organisms since McCarty's data base was primarily freshwater. I would probably expand the safety range for untested species-species in marine waters to 10-100 solely due to the lack of data, unless of course there is recent work that could be used to compute a more accurate range of variability.

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

Answer: If the question refers to WQCTLs then the answer is yes if your referring to PAHs then the answer is for freshwater fishes but not for marine fishes or invertebrates until a comparable data base is developed.

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

Answer: Regarding the WQCTLs they are probably somewhat over conservative but not too much. Regarding the PAHs the CBR is very over protective of invertebrates and questionable for

freshwater fishes since most of the data is for fathead minnows. Nothing in there for trout nor for marine fishes. The solution for PAHs is to develop a larger data base.

Calculations

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: The approach recommended in the Green Book (p 9-8 and 9-9) for estimating total PCBs by summing the individual 22 congeners of concern should be continued as it more accurately represents the PCB concentrations in the samples than by measuring total arochlors. Further, this is the approach used by NOAA and reflects the congeners relevant to environmental abundance, persistence, and most importantly biological importance.

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: Yes the adjustments are appropriate given they provide value added to the decision-making. For example, with dioxins, the residues in Nereis at 28-days were only ~25% of the steady state value achieved after 180 days. If the differences between polychaetes and bivalves occur across a wide range of chemicals within certain log K_{ow} ranges then adjustments could be predicted. Another way to look at the question is to determine the maximum difference between the 28-day tissue concentration and the steady-state value and compare that difference to the variability in replicate bioaccumulation tests. If the difference is less than the variability then one could argue that the additional effort to obtain steady-state values would be lost in the noise. If the difference is $\geq$ 50% as it appears to be in the bivalves then I wouldn't be too concerned but if it was $\geq$ 50% as it appears to be in the polychaetes I'd conduct a more extensive examination.

The application of a 'multiplier' to other listed contaminants should only be considered if there is data to support it.

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: Yes, I thought the description and rationale was well thought out. The only question is that the BaP Toxicity Equivalence estimate relies entirely on data collected in 1980s. Isn't there more recent data that can be used to support the derivation of the equivalence value of 8,021ppb ?

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: The selection of 22 PCB congeners was based upon their toxicology (e.g, potency), and biological importance, bioaccumulation potential, persistence, and presence in the environment. If, using these or some other set of criteria a case can be made for sub-setting the PAHs then it should be done particularly if it provides value added by providing additional predictive power, scientific defensibility, or reduction of uncertainty for the risk manager.

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: Yes, assuming the assumptions, arguments and data used to support them are accurate.

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: Although EPA uses 6.5g/day as their default consumption rate a case can be make for exceptions where subsistence fishing by specific sub-populations are an issue. I can't make a case for increasing the consumption rate for fish coming from that site unless there is specific subsistence fishing there. If the catch is going to the broader market then the 6.5 g/day is fine. I certainly can't make a case for increasing it.

General

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 do think that 'intended use' is appropriate to use for a site when one sets goals for the ecological condition of the site (e.g., community and population endpoints). I'm not sure how to implement the concept is but I do think 'intended use' is relevant.

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

Answer: As stated in the report, this is a first step in the characterization process and one that seems to be adequate for the intended purpose. Second the report recognizes deficiencies in QA/QC and has omitted samples from their estimate as well as spatial and temporal sampling heterogeneity. The use of the Lake Ontario Toxics Management Plan appears sound as do the seven criteria used to make a determination of concern. A decision to remove chemicals should emerge from this review. However, the decision to add chemicals would likely come from the analysis of recent data or other toxic regulatory policies that showed other chemicals are a cause for concern. I noted that the basis for the selection of

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: I don't think it is feasible at this time because to my knowledge there are no models out there that address this issue at the concentrations occurring in the environment. First, the issue of synergistic effects is often a 'red herring' because we don't have enough evidence that this is occurring on a wide scale. Of course lack of evidence is not proof that interactions are not occurring. Further, synergistic effects are generally thought of being additive when in fact they could just as easily be antagonistic but either case is difficult to demonstrate at environmental concentrations. Yes there are laboratory studies that suggest this occurs but the number of compounds and interactions are very limited and in no way capture the scope of the potential problem in the environment. Don't get caught in this trap!

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: This depends on the magnitude of the variability in the data. There is no explanation of the 10X derivation in the report or are you just suggesting using that figure? What makes me nervous is the statistical analyses in Table 1. Here you state that both cadmium and mercury residues in the test sediment are statistically greater than in the reference. Comparing Columns 1 and 3 for cadmium indicate a less than two fold difference in the means for cadmium (0.043ug/Kg vs 0.076ug/Kg) is statistically significant and for mercury it is even less credible - 0.034ugKg and 0.040ug/Kg are claimed to be statistically different. Given the variability of natural samples and the variability of analytical procedures I find these numbers troubling. Likewise for Zn 11.83ug/Kg and 14.34ug/Kg are significantly different. I find it hard to believe the statistics let alone the ecological significance of such differences. So two points are raised by Table 1: how good are the assumptions that have gone into the statistical analyses and what is the potential value of statistical significance relative to biological significance. We have many cases where differences can be statistically significant but be meaningless to the biology. There is rarely a relationship between statistical significance and biological significance Ð purely surreptitious. Statistics is looking at variability, the less variability in a measure then the more power to detect small differences Ð no biological corollary whatsoever. My other concern with the statistics is dealing with 'below detectable' values. What do you choose and why? Many folks say don't make comparisons when you don't have measurable values

with their variability. A second approach is to use the upper limit of detection since you can at least argue with some degree of confidence that the reference sediment is not higher than that value but you have no idea how much lower. On Page 9 it is stated that 'Exceedence of reference values is common where reference values are very low or 'non-detect' as here.' There is no indication of how "non-detects" were handled.

Proposing a 10X exceedence as a more reasonable indicator of potential ecological effects can only be determined if there is evidence to support that hypothesis. Unless analyses are conducted to determine the incremental increase above background where effects occur then selection of 10X is arbitrary. A suggestion might be to attempt to quantify the from sediment contaminant concentrations to tissue residues to ecological effects and in so doing develop a basis for selecting minimally important magnitude differences that are ecologically important. By utilizing the enormous benthic data bases where community structure, sediment chemistry, and toxicity have been measured simultaneously (Chapman's Triad Concept), the data bases where sediment chemistry and bioaccumulation are measured and bioaccumulation and ecological effects are compared one might be able to develop such a relationship to support the 10X or some other factor.

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 assumptions, presented on page 14 pertaining to comparisons of tissue residues from organisms exposed to test sediment with organisms from the vicinity of the remediation site valid for evaluating undesirable effects?

Answer: The decision to preserve the variability of the data sets by using even weighting for all concentrations was appropriate as was the use of the mean as a measure of central tendency. I will assume that standard statistical procedures were used. I think the assumptions for comparing test sediment tissue residues with tissue residues from similar species in the vicinity of the site compliment the comparisons of test sediment tissue residues to residues derived from reference sediments and offer a middle ground for the manager. In other words, the material is better than what is on the site but not as clean as a reference.

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: If the "no further degradation principal" is adopted, then the HARS residues could be used as a baseline. The assumption being that test sediments resulting in benthic residues statistically similar to HARS would present no further degradation. If however, the HARS benthic residues for the test sediments exceed the HARS then one could argue further degradation and reject the use of test sediment. Obviously one question would be how do the HARS site tissue residues compare with reference site tissue residues?