UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
National Center for Environmental Assessment
Washington, DC 20460

NCEA Washington Office (8602))

June 1, 1995

MEMORANDUM

SUBJECT: Review of Proposed Plan for Micro-Exposure Risk Assessment for Upper Hudson River Project by ChemRisk

FROM: Karen A. Hammerstrom
Environmental Scientist

TO: Marian Olsen
Region II
Technical Support Section

Kevin Garrahan asked me to review the above proposal. I have confined my comments to the Monte Carlo assessment and the assumptions about population mobility. Here are the answers to some of your questions:

Is this method statistically and scientifically sound?

This method is acceptable. I believe most of the questions will arise in the evaluation of the input variables. There should be a method for reflecting uncertainty in the input distributions. For example, if a distribution for fish intake is specified, and one of the parameters is an average intake equal to 7 g/day, what is the possible range of values of the average? Could it be as high as 30 g/day for the target population? Could it be higher? By specifying distributions of the parameters of the population distribution, one can generate a series of possible distributions and a range of possible central and high-end exposures. I have an example of such an analysis if you are interested.

How does this method compare to a standard Monte Carlo analysis?

This is a standard Monte Carlo analysis. EPA has always advocated taking into account correlations between input variables and utilizing all available information. One problem with the algorithm is the cutoff after 30 years. The program should run until the point that the incremental risk is negligible. This will depend on how rapidly the PCB concentrations are expected to degrade, how one defines negligible incremental risk, and

how one chooses to define the exposed population (see discussion below). The framework of the analysis should improve accuracy because it takes correlations between input variables into account. The assumptions and data sets used to evaluate the input variables are more likely to have an impact than the algorithm itself.

How do the results of the analysis compare to the standard RME calculation?

In strictly numerical terms, this also depends on how the input variables are evaluated. The RME contains an element of uncertainty through incorporating the UCL of the average exposure concentration. Using the UCL has little effect on the result if the exposure concentration data are adequate. If there are few samples with widely varying concentrations, the UCL may be very much higher than the mean concentration. This is more a reflection of inadequate site characterization, than of built-in conservatism in the assessment.

In my opinion, a number comparable to the RME would be around the 95% upper confidence limit on the 99th %ile exposure/dose level. However, those who compare the RME to Monte Carlo results usually assume that the RME is intended to represent something like the "most likely value" (say the median or mean) of the 95th %ile exposure/dose, which is determined by taking the average or median value of all 95th %ile values generated in the outer loop run of the Monte Carlo analysis. The latter approach could arguably leave 5% of the population exposed above the concern level. The counter argument is that conservatism is built into the cancer slope factors and RfDs.

Are you familiar with any other sites, etc. that used this approach?

I saw a Region 4 assessment in 1993 that was very similar, but I don't know the site. Elmer Akin sent it to Bill Wood of the Risk Assessment Forum for review. It involved fish ingestion in a reservoir, and went through the procedure of randomly selecting individuals and assigning body weights and intake rates by age.

What are the pitfalls associated with this approach? What are the benefits?

Pitfalls - It's much more difficult to produce a supportable, intelligible assessment. Very often the authors do not have the expertise to set up such an assessment, and they have major problems evaluating input variables. Review of Monte Carlo assessments is very labor intensive. I spent 120 hours reviewing the East Fork Poplar Creek assessment for Region 4 and 80 hours reviewing the Yeoman Creek assessment for Region 5, and only completed partial reviews. In both these assessments, there were problems with the evaluation of practically every input variable.

My recommendation has been that the authors focus on one or two important pathways and discuss evaluation of input variables with EPA as they proceed. This approach will allow the organizations producing the analyses to focus more expertise on evaluating the input variables, will decrease the number of input variables that EPA has to review, and will illuminate many of the potential issues long before the Monte Carlo assessment is presented. Since this is what you are doing, I believe you have a much

greater opportunity for success in this case.

Benefits - Monte Carlo is a tool for combining input variables. Properly used, it is an excellent way to show exposure, dose and risk distributions as well as to reflect the uncertainly in the distributions and correlations between input variables. The method is vastly superior to the point estimate in cases where the data base supports evaluation of the variables. In cases where there are little data for most input variables, I prefer the point estimate, however. In the Hudson River case, a MC analysis is appropriate.

Are all of the parameters random probabilities? Is there a potential to propagate errors in each component of the analysis?

I'm not sure I understand this question. I don't see any provision for propagating uncertainty in the values of the parameters of the population distributions. This is discussed further below.

What are the minimum data sets that will be needed to run this model? Would this model run using surrogate data -- what are the pitfalls?

I believe there are sufficient data to run the model, but there are likely to be some disagreements between ChemRisk and EPA as to the evaluation of some of the parameters. I'm not sure what you mean by surrogate data. Do you mean, for example, using intake values for some other population to represent those who fish in the Hudson River. If so, using surrogate data is acceptable as long as you try to pick the most representative data set and adequately assess the uncertainty in that data set.

What is the minimum number of runs? Hardware and software requirements?

The total number of iterations should probably be on the order of 1 million. Five thousand runs of the inner loop (population variability) are sufficient, but they can probably get by with fewer. However, they should also do multiple runs of an outer loop which evaluates uncertainty in the parameters of the inner loop distributions. I'm not sure how many outer loop iterations would be required. I don't think 5,000 would be necessary. Tim Barry could give you a better idea, because he did this in the radon assessment. He can also tell you about software and time requirements. I believe these runs can be made on a 486 PC, but special programming is required.

Issues identified in cursory review.

These issues all relate to evaluating the "duration" term. As a default, many Agency assessments have used 30 years, which I believe is the 90th percentile value of the time people reported having lived in their current residences in the Census. It's difficult to understand what Chem Risk is proposing without having the paper "*Estimating Exposure Duration for the Hudson River Risk Assessment.*" (This paper is apparently not in the list of references.) I believe Chem Risk is proposing the following: The goal is to produce a continuous distribution reflecting the probabilities of various exposure durations for the defined population. The algorithm approximates this continuous distribution by a year-by-year construction of the relevant activities of 5,000

hypothetical individuals. The exposed population seems to be those who fish in the Hudson River, rather than those who consume fish from the Hudson River. Exposure of an individual is assumed to end after 30 years or when the individual moves out of some specified area, dies, or stops fishing. The exposure does not include individuals who move into or are born into the study area after Year 1. The Chem Risk approach could be thought of as an assessment of a highly exposed subpopulation (anglers fishing in Year 1) of the total population (anyone who consumes Hudson River fish during the duration of the exposure period). Chem Risk's approach should produce somewhat higher average and high-end individual exposure estimates than an assessment of the total exposed population. This is not an overestimate, but rather an estimate for a subpopulation.

I recommend two changes in the approach -- 1)moving back to the area after moving away and resumption of fishing after stopping be considered and 2) that the analysis continue for longer than 30 years unless incremental exposures are negligible after 30 years.

There is no reason to arbitrarily stop the analysis after 30 years. The algorithm should be continued until the incremental exposure is negligible. This could be after 10 years, 50 years, or even 1,000 years if the total exposed population of anyone who ever eats PCB-containing fish is assessed.

Other comments

Page 7, paragraph 1: What is the argument for ignoring past exposure?

Page 7, nos. 1 and 2: If the microexposure model is used, the analysis should not be cut off after 30 years.

Page 9, paragraph 2: The probability calculation cited does not apply in this case. The likelihood of an individual's receiving a dose greater than the RME depends on the relative standard deviations and shapes of the distributions of the input variables and the functional form of the model. If the RME is calculated using a multiplicative model with three variables set equal to their 95th percentile values, the corresponding dose value in the Monte Carlo analysis will be greater than or equal to the 95th percentile value. If two variables have low relative standard deviations compared to the third, the exposure/dose distribution will correspond closely to the distribution of the third variable. The other two variables will behave almost like constants.

cc: Tim Barry
Kevin Garrahan
Paul White
Jackie Moya