PEER REVIEW OF EPA'S HAZARDOUS WASTE IDENTIFICATION RULE RISK ASSESSMENT MODEL

Conceptual Approach to Establish Interim Human Health Benchmarks

Prepared for:
David Bartenfelder
Office of Solid Waste
U.S. Environmental Protection Agency
2800 Crystal Drive
Arlington, VA 22202

EPA Contract No. 68-W-99-001
Work Assignment No. 17

Prepared by:
Eastern Research Group, Inc.
2200 Wilson Boulevard, Suite 400
Arlington, VA 22201-3324

December 17, 1999

NOTE

This report was prepared by Eastern Research Group, Inc. (ERG), an EPA contractor, under Contract Number 68-W-99-001. The report presents comments provided by peer reviewers on the *Conceptual Approach to Establish Interim Human Health Benchmarks* document that is part of EPA's Hazardous Waste Identification Rule risk assessments.

The comments presented in this report have been compiled by topic and by individual peer reviewer. As EPA requested, this report provides the peer review comments exactly as they were submitted to ERG. Also attached are the original comments submitted by each individual reviewer.

Charge to Reviewers

Conceptual Approach to Establish Interim Human Health Benchmarks

BACKGROUND

The purpose of this report is to describe the methodology that the Office of Solid Waste (OSW) is proposing to support development of interim human health benchmarks for constituents lacking EPA benchmarks. These benchmarks are intended to be used by OSW in the Hazardous Waste Identification Rule (HWIR) risk analysis and other OSW risk assessments. The HWIR risk analysis is being conducted to help OSW establish chemical-specific waste concentration limits, referred to as exemption criteria, that would be protective of human health and the environment based on selected Agency protection measures.

There are over 450 constituents under consideration for HWIR, many of which currently lack any Agency-derived human health benchmark (i.e., reference dose/concentration, cancer slope factor, or unit inhalation factor). Therefore, it is not possible to develop exemption criteria for these constituents, likely limiting the number of potentially hazardous waste streams eligible for exemption. Because the resources required to develop benchmarks using the traditional approach exceed the time and budget constraints for the HWIR analysis, OSW directed the Research Triangle Institute (RTI) to develop a practical and defensible methodology to establish interim benchmarks in an expedited manner. The interim benchmarks are conservative in the sense that they are based on methods and data sufficient to support a high level of confidence that exemption criteria are equal to, or below, criteria that might be developed using "standard" EPA sources such as the Integrated Risk Information System (IRIS).

The overall goal for the peer review process is to ascertain whether the proposed methodology constitutes an appropriate tool to develop interim health benchmarks using existing information in a timely manner for the HWIR risk analysis (e.g., less than two years). In particular, OSW is interested in receiving comments suggesting: (1) areas of improvement for the methodology such as data sources or regression techniques, and (2) alternative approaches for interim benchmark development not described in the report.

OSW requests that your review focus on three areas: (1) general issues involving the overall strategy of the methodology (i.e., does the methodology achieve the stated goals?), (2) issues relevant to the development of interim cancer benchmarks, and (3) issues relevant to the development of interim noncancer benchmarks. It should be noted, however, that the charge issues are not intended to be prescriptive; OSW welcomes comments on any technical issues associated with an approach or the data relied on to develop interim benchmarks.

GENERAL ISSUES FOR REVIEW

1. OSW is aware that the use of alternate methods to develop health benchmarks is contentious. Do you feel that the major sources cited (e.g., Hoover et al., 1995) in this report have gained sufficient acceptance in the risk assessment community for application in a regulatory context?

2. Are there alternative methods not presented in this report that you would recommend for use in deriving interim health benchmarks?

3. Do you support the logic illustrated in the Figure 3-1 hierarchy for development of interim benchmarks? If not, what changes would your recommend to strengthen the hierarchy?

4. For constituents lacking dose-response data, limited use of chemical analogs has been proposed. Examples used by EPA include toxicity equivalency factors (TEFs) developed for dioxins and PAHs, as well as the use of a sulfate or salt form of a compound as a surrogate. Do you believe that, given the goal of the expedited methodology, these approaches are scientifically defensible or should be expanded to include additional techniques and constituent groups?

5. Following interim benchmark development by one or more of the expedited methods, the methodology suggests that an expedited review process be conducted. At this point, the expedited review has only been conceptualized. Since this review could be accomplished in a wide variety of venues, please provide recommendations for an expedited review process that you feel would meet reasonable standards for scientific rigor.

CANCER ISSUES FOR REVIEW

1. For constituents for which one or more dose-response carcinogenicity studies exist, the linearized multistage model will be used to estimate the CSF. The uncertainty in the interim benchmark will be characterized by considering the additional uncertainty introduced by the use of the expedited process. The expedited cancer benchmarks will be adjusted by an R factor estimated from a distribution of ratios of cancer benchmarks developed by both an expedited and traditional approach. Do you agree with the proposed use of the R factor to account for the uncertainty inherent in an expedited approach and, if not, please recommend alternatives for consideration by OSW.

2. The relationship between TD_{50}s and CSFs (and/or URFs) demonstrated through regression analyses is proposed for use in developing cancer benchmarks (i.e., the CSF and/or URF is estimated as the lower 95% confidence interval from the TD_{50}). Do you agree that this is an appropriate technique to derive interim cancer toxicity values? If not, do you have

recommendations to improve upon this technique or can you suggest other methods that you consider to be more appropriate for this application?

3. The use of the TD_{50}/CSF relationship provides a much quicker, less resource-intensive method of cancer benchmark derivation than using the linearized multistage model (for constituents that have dose-response data available). Do you think that these advantages are sufficient to justify a change in the methodology such that this method is used in lieu of the linearized multistage model for constituents that have quantitative cancer data available?

NONCANCER ISSUES FOR REVIEW

1. For constituents possessing one or more dose-response noncancer studies, the data will be examined to identify the most appropriate NOAEL or LOAEL. To account for the uncertainty in using a limited data set, probability distributions of uncertainty factors (in Baird et al., 1996; Evans and Baird, 1998; Swartout et al., 1998) could be used to define quantitative measures of confidence for the interim noncancer toxicity values. The resulting adjustment factors serve a similar function to the uncertainty factors used in the traditional approach. Do you believe that the use of probability distributions is an appropriate alternative to the traditional uncertainty factors given the limitations of the supporting data set? If not, would you recommend specific changes to how the probability distributions are used or can you suggest an alternative approach that is more technically defensible?

2. Lacking the requisite dose-response data, or information on a chemical analog, the next option for estimating a noncancer inhalation benchmark is based on the regression of threshold limit values (TLVs) against RfCs. This regression has been performed and correlation coefficients developed corresponding to specified confidence limits. Do you consider this an appropriate use of these data and can you suggest improvements for this approach?

3. Representative distributions have been developed for current RfDs and RfCs that can be used to establish interim benchmarks for data-poor (i.e., no appropriate studies available) constituents at a 95% confidence level. Naturally, there are disadvantages to applying a single value to a group of data-poor chemical constituents (e.g., likely to be overly conservative for some constituents). From a technical standpoint, do you agree that this approach achieves the desired goal of the methodology, namely, that the interim benchmark is equal to, or more conservative than, the traditional benchmark at a specific level of confidence? If you disagree, can you recommend alternative approaches (or improvements) to address these data gaps? Note: Refinements to the distributions based on chemical class are discussed below.

4. The representative RfD and RfC distributions may be modified and improved by subdividing the universe of chemical constituents into classes based on physical and chemical characteristics (i.e., use of SAR information). The methodology proposes a number of distributions for use in estimating interim benchmarks as a function of chemical class. Using the scheme, the interim benchmark is chosen from the class distribution rather than distribution of all constituents. Do

you think that the use of SAR in this scheme is an improvement to the use of distributions given the uncertainty associated with chemical/physical properties and toxicity mechanism of action? Do you agree with the chemical classes developed to refine the distributions? Can you recommend changes to the chemical class scheme that would improve the defensibility of this approach? Can you recommend criteria other than chemical class that might be used to refine the distributions?

Reviewer Comments Summary Report for the
Conceptual Approach to Establish Interim Human Health Benchmarks

GENERAL COMMENTS

Dr. Hoover:

Initial Comments

The overall goal of the methodology is clear and provides the underpinning for decisions made throughout the document. I am in support of the use of expedited approaches, which have a two-fold benefit as has been pointed out by other authors: 1. Allow for the regulation of previously unregulated compounds; 2. Encourage the testing necessary to develop a less conservative benchmark. The application of the expedited philosophy should be strengthened and made more consistent, however; detailed comments in this regard are provided in later sections.

Abbreviations

Abbreviations are consistent with those used in the charge to reviewers and the document.

Highlighted Issues

The main issues identified in the review are highlighted below:

A more detailed summary of the context for this document (i.e., proposed HWIR) along with appropriate citations should be included in the Executive Summary and Introduction. Some of the background information is already available in the charge that was provided to reviewers and can easily be included in the document. Appending the entire list of the constituents known to be of concern under the HWIR would also be useful to reviewers and users of the document.

The approach to deriving interim cancer benchmarks should be revised and simplified to achieve the goals of an expedited method (see specific comments in later sections and a summary of a proposed revision beginning on page 16).

The document does not provide enough detail on the primary method for expediting the derivation of noncancer benchmarks. The approach to limiting the data set under consideration must be specified and/or alternative options considered (see specific comments in later sections and a summary of a proposed revision beginning on page 16).

I strongly support the use of more sophisticated techniques to derive noncancer benchmarks, including, for example, the use of benchmark dose modeling, consideration of uncertainty in dose-response model and use of probabilistic uncertainty factors. However, even though the science of evaluating uncertainty in toxicity benchmarks has progressed considerably, these techniques have not been widely applied in a regulatory context (Roberts, 1999) and therefore may not be compatible with the goals of the expedited philosophy. The field of probabilistic uncertainty factors in particular is still

evolving.

When using statistical methods to derive RfCs/RfDs in the absence of specific data on a chemical, the chemical should be evaluated for potentially unusual properties that could affect the toxicity. It would also be useful to conduct a "real world" check on the toxicity value derived using these methods (see page 10 for more details).

The document would benefit greatly from the extensive use of illustrative examples.

Dr. Juberg:

If it is not obvious in the following page-by-page comments, it will become rapidly apparent that this reviewer believes there is an excessive emphasis on statistical treatment of data related to development of this interim approach and conversely, not enough biological plausibility and toxicological consideration given to the methodology. The authors will need to defend this position and explain to the users of such an approach, why use of statistics, R factors, distributions, and log functions are so critical at the exclusion of other more biologically meaningful parameters.

This approach and proposed methodology may be used for a variety of purposes associated with OSW activities and responsibilities and it would be very helpful to have some background in the Draft document pertaining to such, as this will help users and reviewers in knowing what types of exposure scenarios humans might encounter. This type of information is critical and relevant in the subsequent development of appropriate health benchmarks. For example, in setting an occupational standard, the route of exposure is frequently inhalation and the temporal consideration is usually 8-10 hrs/day. For purposes of better developing the proposed approach in the Draft, it is imperative to identify the range and types of exposures that humans may encounter – Will exposures most likely be through inhalation, oral ingestion, or dermal contact? Will exposures be transient in nature or lifetime chronic exposures? If these interim benchmarks are used for exit criteria associated with an industrial facility, it is unlikely that humans will ever have chronic exposures to these various constituents. The approach that is developed and used needs to mirror, to the extent possible, those most likely anticipated human exposures and scenarios. In this regard, the methodology needs to anticipate and forecast the exposures (inhalation, oral, dermal) and durations (minutes, hours, lifetime). Such will be critical in a final analysis as to how conservative each interim benchmark needs to be. Additionally, the OSW should provide some background information on its various regulatory programs and activities, such that users of this methodology will know the basis and impetus for development of such an approach.

A rather large portion of the Draft pertains to the Appendices illustrating the statistical treatment (e.g., log RfDs, etc.) of all the constituents under potential review and the appendices appear to be disconnected from the text. In other words, there needs to be additional explanation or linkage in the text as to what and why the appendices are included. There should be explanation within the appendices themselves as to how the statistical treatment of the constituents is related to the proposed

benchmark establishment process. Give examples. Finally, similar to a previously mentioned comment, the primary focus should be on the text and on explanation of the process; minimal treatment should be given to exhaustive presentation of the constituents (in the appendices) at this stage. The detail given to the appendices would be have been better served by showing examples of how the process may or may not work.

There needs to be some mechanism within the proposed approach for provision of external peer review and the ability of non-Agency scientists to participate, either directly or indirectly, in the development and adoption of interim benchmark values. Historically, the Agency has limited participation from external parties, yet there is positive benefit to having multiple stakeholders and interested parties participate in this process.

There needs to be some provision for "field-testing" or validating this approach prior to any further development or promulgation. While the approach has some merit as a screening tool, there needs to be evaluation and validation of this approach which could be accomplished by taking several different constituents and developing health benchmarks based on the contents of this document. If the resulting values appear far below either natural, background, or detectable ambient levels, then some adjustments within the methodology are required. Clearly, however, there needs to be some evaluation and validation of "test case constituents" before the methodology is further developed.

It is somewhat troublesome that a primary goal of the approach is to insure that any developed interim benchmark is more conservative than traditional health benchmarks. This puts the onus of benchmark development on preserving a policy goal and not allowing the science and knowledge drive the process. This will become quite apparent to users (and abusers) of this approach and may lead to a less than desired acceptance of the methodology by the regulated community. This reviewer would advocate using sound scientific principles and data and letting the "chips fall where they may." As we have learned over the years, most regulatory health benchmarks are very conservative by their nature and intent, and thus, here is an opportunity to shift the policy to one that is not so conservative and more in line with the scientific evidence.

This approach and methodology will be of limited use to most risk assessors who have not had extensive statistical or probabilistic training. Much of the discussion is beyond what many scientists can understand without additional training and education, and given the fact that this is an interim benchmark approach, the level of statistical sophistication is excessive.

Dr. Zeise: The overall strategy to establish benchmarks through an expedited strategy is a good one and the Agency has been thoughtful in its approach to the matter. However some of the techniques outlined required adjustments to reduce the uncertainty and bias.

GENERAL ISSUES FOR REVIEW

***Charge 1**:* *OSW is aware that the use of alternate methods to develop health benchmarks is contentious. Do you feel that the major sources cited (e.g., Hoover et al., 1995) in this report have gained sufficient acceptance in the risk assessment community for application in a regulatory context?*

Dr. Hoover: After detailed review of the document it is apparent that the proposed approach to deriving interim cancer benchmarks relies only in a minor way on Hoover et al. Thus regulatory acceptance of the Hoover et al. approach does not seem particularly relevant to the currently proposed approach (comments on revising the approach are provided later). That being said, in answer to this question Hoover et al. was used as the basis for a lower tier of cancer potency factors which have been promulgated as regulatory values under Proposition 65. Thus the values are certainly accepted in that regulatory context. I conducted a search in the scientific literature for papers that have cited Hoover et al. but located only Cranor (1995). I also looked for references to expedited approaches on the Internet but did not locate any. I am not aware of other state or federal agencies that have applied the Hoover et al. approach or the cancer potency factors reported therein in a regulatory context.

With regard to sources such as Swartout et al. and Baird et al., for example, the field of applying probabilistic techniques to establishing benchmarks is evolving and new papers are being published on an ongoing basis. I conducted a citation search on Swartout et al. 1998 and Baird et al. 1996 and found a number of more recent publications (listed in the reference section at the end of my comments). To my knowledge, the use of probabilistic uncertainty factors has not been accepted in a regulatory context as yet.

In terms of sources regarding TEFs, these are well accepted and have been applied in numerous regulatory circumstances.

As a final note, I do not think that lack of precedent should necessarily influence what is done for the purposes of HWIR. The expedited approach applied by Hoover et al. under Proposition 65 had no precedent but was accepted in that regulatory context. The expedited approach selected by OSW should meet the regulatory requirements and goals of the HWIR.

Dr. Juberg: This reviewer believes that the risk assessment community is not sufficiently proficient and knowledgeable in the uses and limitations of some of the major approaches (e.g., R factor) in this Draft to warrant application in a regulatory context (without some additional discussion). Some of the other citations (Baird, Swartout) are more common within this context.

Dr. Pascoe: As pointed out in the report, alternative methods that are considered less rigorous than standard acceptable methods will be unacceptable to a portion of the scientific community. As that may be, the overall concept of deriving interim benchmarks and the methods proposed in the report are defensible within the context of their intended use and purpose. I believe that one of the larger problems with the overall concept is not so much the technical approach to deriving the benchmarks, but the subsequent use of the benchmarks in applications they are not intended for. It

is common for both regulators and the regulated community to rely on rapid means or readily available numerical benchmarks for assessing risks from public exposure to chemicals. This has often been the case in evaluating contaminated properties and other uses of risk assessment outside of federal agency oversight. Any final product or benchmark developed under the proposed expedited approach must have cautionary language against the application of the benchmarks to unintended uses, such as cleanup of hazardous waste sites.

The major sources cited in the report are from highly regarded peer review journals and are sufficient for the intended purposes of the approach. I am not aware of other sources that have undergone a peer-review process.

Dr. Zeise: Expedited Cancer Potency Development

Values generated by the expedited technique of Hoover et al. (1995) are being used in various regulatory programs in state government risk assessments in California. Expedited values have been employed in air toxics "hot spots" assessments, hazardous waste evaluations, and Proposition 65 enforcement. Thus they have achieved regulatory acceptance. It is important to acknowledge that different risk assessment needs require different levels of evidence, and that the expedited values are primarily being used in "right to know" context (Proposition 65 and Toxic Hot Spots). When the regulatory impacts are more far reaching, such as widespread regulation (e.g., of an air toxic contaminant), more labor intensive methods are employed. Also, risk assessment structure for establishing potencies in California under programs utilizing expedited values is hierarchical, with differing levels of review and appeal, depending on the level of scientific investigation involved in establishing the number (see e.g., California Code of Regulations, Section 12705).

It appears that the risk assessment context under consideration for the application of an expedited approach would benefit from the use of such an approach. That being said, there are important technical differences between the approach described in the document and that of Hoover et al., so answering the charge question in the affirmative can not be seen as an indication that the approach laid out in the OSW document should be used in a regulatory context.

Charge 2: *Are there alternative methods not presented in this report that you would recommend for use in deriving interim health benchmarks?*

Dr. Hoover: The document discusses the major options. I would expand the reference list to include more extensive citation of the literature on petroleum hydrocarbons. The surrogate approach is widely applied in that field. The use of surrogates in that field is also a recognition of the value of an expedited approach, as the need for doing something rather than nothing in the face of large data gaps for petroleum hydrocarbons motivates the use of surrogates. With regard to other possible methods in the literature, I would suggest conducting a current literature search to check for suitable references as the field of risk assessment is a rapidly evolving one.

Dr. Juberg: Presently, there are various methods and approaches for prioritizing chemicals in terms of toxicity (and bioaccumulative capacity and persistence within the environment), but there are fewer widely known, used, and accepted methods for deriving health benchmarks. The current approach described in this Draft could be appropriately used as a first tier, screening method for purposes of putting some type of framework around hundreds of constituents. However, and very importantly, a chemical-specific subsequent review or check must be conducted to determine whether the approach described in this Draft has appropriately characterized the chemical's toxicological properties. For example, if the Draft approach were used and an interim benchmark determined for a constituent, and yet a followup check determined that the value was lower than current natural background levels, then the methodology and interim benchmark would have to be revisited as the approach would have been ineffective in its intent and purpose. If many or all values developed within this approach turn out to be very low values, those approaching background values, then the approach will not be effective as a screening tool.

It is also recommended that following development of screening values, there be some mechanism for conducting a site-specific review, based on the intent and interests of the petitioning party. In other words, if an industry or owner of a property could demonstrate through risk assessment or other viable techniques, that there are no human exposure or receptor pathways, then the developed interim benchmark may be increased (less conservatism) to a mutually agreeable level.

Dr. Pascoe: Of alternative approaches to developing interim benchmarks that I have encountered, none are recommended for inclusion in the report. I am not aware of alternative approaches that are more appropriate or more defensible than those presently in the report.

Dr. Zeise: There are other toxicity based approaches that could be considered (e.g., Zeise et al., 1984; Travis et al.). There will be less general consensus about the use of these approaches, so they should be employed only if the Agency believes it can withstand the criticism for doing so. Such approaches would enable the coverage of a greater number of chemicals with carcinogenic activity than is provided for in the current scheme.

Charge 3: *Do you support the logic illustrated in the Figure 3-1 hierarchy for development of interim benchmarks? If not, what changes would your recommend to strengthen the hierarchy?*

Dr. Hoover: First, the diagram on page vi illustrates a different hierarchy than Figure 3-1. Most of the descriptions of the process provided in the document are consistent with the figure on page vi (the one exception is discussed below). For example, the document indicates that for chemicals identified as carcinogens, the first step would be to identify whether a cancer benchmark study is available, and if not, then determine if the chemical analog criteria are met. Figure vi is consistent with this. Figure 3-1 is not. Figure 3-1 appears to show that chemical analogs would only be considered if the TD50/CSF regression cannot be done. The meaning of the double arrow in

Figure 3-1 connecting the "Cancer Benchmark Study" box with the "TD50/CSF Regression" box is also unclear, though I assume it means these two options are considered to be at an equivalent level. Based on the document, however, it would appear that the first step after identification of a carcinogens should be to evaluate whether a cancer benchmark study is available, if not then consider chemical analogs and if no chemical analog is appropriate then consider use of TD50/CSF regression. On page 16 in the first sentence of Section 3.1.3, however, the process appears to be described differently: "For some constituents that do not have a **cancer benchmark** [as opposed to a cancer benchmark **study** which is the phrasing in the rest of the document] there may be dose-response data that have been used to develop an alternative carcinogenic potency estimate... This statement would appear to be consistent with the logic in Figure 3-1. Once the method is revised and finalized, I suggest preparing one figure that reflects the appropriate hierarchy.

Second, the diagram does not explain some of the key decision points. For example, in the portions of the document describing interim cancer and noncancer benchmark development, the concept of evaluating/analyzing only a "subset" of studies is mentioned. There should be a diagram showing what logic will be used in identifying the subset of studies, if this aspect of the proposed approach is retained.

Third, I have concerns about the first step in the hierarchy, in which the cancer designation available from US EPA, IARC or DHHS is used to identify carcinogens. Given the lengthy review time required to make a cancer designation by any of the three named agencies, this first step in part defeats the expedited philosophy taken in the current document. For example, suppose an NTP bioassay is available, in which clear evidence of carcinogenicity in animals is supported by statistically significant increases in tumors in two species. If the NTP bioassay is relatively new, the EPA, IARC and DHHS designations are unlikely to be available, yet the likelihood that this chemical would be considered a potential human carcinogen once the reviews have been completed by these agencies is high. I suggest providing for some mechanism to make an expedited determination of carcinogenicity, rather than setting a noncancer benchmark in a clearly inappropriate case. Another option would be to expand the list of designations to include chemicals known the state of California to cause cancer under Proposition 65, because that is a more comprehensive list which includes chemicals identified as carcinogens by EPA, IARC, NTP, FDA, and NIOSH, as well as the state's qualified experts.

I also note that the use of these designations is not consistent in the document. On page vii, chemicals in the EPA A, B, C or "known/likely" classifications are considered carcinogens, whereas on page 10 only the A, B and "known/likely" categories are named. Given the fact that the "C" classification refers to a "possible" human carcinogen I assume that the authors intended to include only the "A, B and known/likely" classifications. This raises an interesting point, however, with regard to using expedited approaches to encourage testing for the purpose of filling data gaps. It would be beneficial to include as carcinogens chemicals that fall into the categories that designate "possible" carcinogens or chemicals with "limited evidence" of carcinogenicity (such as the "C" classification from US EPA and the "3" classification from IARC) as a means of

encouraging the completion of sufficient studies to properly classify these chemicals as to their carcinogenicity. Chemicals falling into these categories would need to be screened first, to confirm that the reason for the designation is that insufficient/limited data are available on carcinogenicity. Certain chemicals with these classifications may actually have a sufficient data set that provides only limited evidence of carcinogenic effects. In these cases, rather than treating the chemicals as carcinogens, an additional safety factor could be applied to an interim noncancer benchmark to address the limited carcinogenic concern. In other cases where the classification is based only on a limited data set the chemical could be classified as a carcinogen and an interim cancer benchmark derived using the limited data. Alternatively, all chemicals in these classes (e.g., EPA "C", IARC "3") could be treated in the same way, that is to derive an interim noncancer benchmark and apply an additional safety factor to protect for carcinogenic concerns.

Fourth, the logic of the derivation of the interim cancer benchmark is problematic based on how it is generally described in the document. The approach hinges on whether or not a cancer benchmark study is available, with other options to follow if no adequate benchmark study can be identified. One of the secondary options is to use a TD50 to derive a CSF based on a regression. If a TD50 is available, then a cancer benchmark study must necessarily also be available, meaning there is no need to use the TD50 to derive a cancer potency. True, the study may not meet the criteria laid out in this document, but application of the linearized multistage model to derive a CSF is still possible. I suggest that applying the LMS model to a limited study would be more acceptable and closer to results expected under the traditional method than use of the TD50/CSF regression would be. The decision taken by Hoover et al. to use the LMS model rather than using the relationship between TD50 and CSF to derive potencies was in part a result of feedback from the scientific peer review of the method.

Dr. Juberg: Two major changes or proposed additions. (1) Following development of the interim benchmark, there needs to be some provision and mechanism for allowance of external peer-review and this should be included in the diagram. (2) Also, there needs to be some flexibility and chemical-specific provisions for what constitutes a carcinogen. Our knowledge in cancer biology has advanced sufficiently (although the major cancer classifying mechanisms such as IARC, IRIS, etc. have not always kept sufficiently up to date in terms of classification) that we now know there exist certain species-, mechanism-, or chemical-specific factors that preclude the direct extrapolation of animal cancer results to humans. Just because a chemical has been demonstrated to be positive in a rodent cancer bioassay and is labeled as a probable or possible human carcinogen is not sufficient cause for moving it through the hierarchy as proposed and setting an interim cancer benchmark. Half of all chemicals tested in standard high-dose animal cancer bioassays, whether occurring naturally or produced synthetically are "carcinogens"; there are high-dose effects in rodent cancer tests that are not relevant to low-dose human exposures and which contribute to the high proportion of chemicals that test positive (Ames and Gold, 1997). This reviewer believes that development of interim cancer benchmarks needs to be evaluated very carefully so we do not erroneously develop such for a chemical that truly has no human carcinogenic potential.

Dr. Pascoe: In general, I agree with the logic of the approach hierarchy as illustrated in Figure 3-1; however, see my comments on the use of TD50s for deriving CSFs. The multistage cancer dose model should be given priority, but following an initial identification of a carcinogen from the database of TD50s.

Dr. Zeise: Selection of Chemicals for Consideration

The figure gives no indication of how chemicals will be selected to enter the identification scheme provided in the figure. Of the myriad of chemicals of potential concern, how will the Agency decide which ones to focus on? Will it be done on the basis of availability of toxicity information? Structure activity? What quality assurance will be in place to ensure that those of greatest risk potential are being addressed first? Some discussion of this issue would be helpful, and if possible, it should be addressed in Figure 3-1.

Carcinogen Identification

The first step in Figure 3-1, how agents are identified as a carcinogen or not, should be reconsidered. The main problem with this approach is the limited number of chemicals that can be addressed by it. The approach is being applied to a wide variety of agents, many of which may not have been adequately tested for carcinogenic activity or undergone the labor intensive problem of carcinogen classification by the named authoritative bodies. The number of chemicals which have been adequately tested in animal bioassays and reviewed formally as part of carcinogen classification may be a very small fraction of the number of chemicals potentially of concern to OSW. Thus, the process would benefit from an expedited hazard identification component. While a comprehensive approach to this problem would be hard to develop, receive regulatory acceptance and implement in the near term, some relatively modest improvements could be made and implemented under the current scheme. Meanwhile, the Agency should, in a long term effort, work on developing a more comprehensive approach. In this regard collaboration with scientists in the Agency's premanufacturing notice program, and consideration of approaches employed in the pharmaceutical and pesticide product development should be beneficial.

The initial carcinogen identification question asked should not be whether or not some body has identified an agent as a carcinogen but the extent to which the agent may pose a carcinogenic hazard. Findings of authoritative bodies such as IARC, EPA IRIS and National Toxicology Program (NTP within DHHS) will help answer this question, but if the chemical has not been reviewed or has been reviewed some time ago, its carcinogenic potential should still be assessed. First, at the simplest level, if two organic salts dissociate into the same active ionic form, and one is listed by one of the three authoritative bodies but the second is not, the Agency should treat the second as a carcinogen. Second chemicals that are metabolized to known/likely carcinogens should be treated as carcinogens. Third, and more difficult, if there are clear structure activity relationships *and* the chemical fits within a class of chemicals with known carcinogenic activity (e.g., nitrosamine, nitrosourea) then it may be appropriate to include it in the group of chemicals with likely carcinogenic activity. Fourth, if a chemical has been studied in bioassay and

clearly tests clearly positive, or has strong positive mechanistic information and suggestive information from the bioassay, it may also be appropriate to treat it as a carcinogen. Fifth, if an authoritative institution besides the three named (e.g., National Institute of Occupational Safety and Health) has identified the chemical as possessing carcinogenic activity or the chemical is on California's Proposition 65 list, it could be treated as such. The existing scheme could be expanded to include these approaches.

Finally, one could investigate the application of a more general approach for identifying carcinogens based upon structure activity and other considerations, but this would take time to develop and could not be included in the scheme for some time. It nonetheless remains problematic to identify chemicals as not possessing carcinogenic activity because of paucity of information. Practical approaches may be developed that are not as data intensive as those of the current classification schemes used by the Agency, IARC and NTP. Those employed in drug and pesticide development could be considered and expanded upon. Ultimately it may be beneficial at some future point to move a probablistic scheme.

A second problem with the scheme proposed is that it may be uneven in carcinogenicity classification. The Executive Summary indicates that EPA group C agents will be treated as carcinogens, but for IARC, those in groups 1 and 2 would be included. Group 3 IARC evaluations finding limited evidence in animals are roughly equivalent to EPA's group C classifications. It seems appropriate to include chemicals with limited evidence as possessing some carcinogenic activity, but some adjustments may be warranted in quantifying their activity. In the case where a chemical has been tested to a limited extent (e.g., limited pathological evaluation, single study) but positive finding resulted and where mechanistic studies suggest a genotoxic mechanism, the application of the approach could be the same as for chemicals with more evidence (e.g., IARC 2B). In the case where the chemical has been extensively tested and found to have weak activity in a few test, and mechanistic information is not impressive, a less stringent approach may be in order. For this case, it may be appropriate in certain cases to apply a standard non-cancer approach to the bioassay dose response data and include an extra uncertainty factor. A benchmark dose would be preferable to a NOAEL as the point of departure for this case. Non-cancer endpoints should also be evaluated to determine if they produce a lower benchmark. Alternatively, as is sometimes done by the Agency in its drinking water program, the benchmark may be set on non-cancer findings, with an extra uncertainty factor applied to account for possible carcinogenicity. This is less preferable but has the advantage of being already in use within the Agency.

Cancer dose response evaluation

The scheme does not indicate how it will address agents which are identified as having carcinogenic activity but do not have data sets meeting the selection criteria (e.g., too few dose groups). The document suggests use of an expanded structure activity approach, and exploration of such techniques for this use are encouraged. Another possibility is to include a toxicity based scheme to estimate carcinogenic potency (e.g., Zeise et al., 1984). It may be that some hybrid of

these two approaches would lead to better estimates. This should be explored. It would greatly increase the coverage.

A major concern regarding the approach laid out in Figure 3-1 is the inclusion of a separate TD50 carcinogenicity measure. If a TD50 exists in the Gold et al. database then dose response data, tabulated in that database, are available for cancer dose response (multistage) modeling, and should be considered for use. There is no reason to rely on a TD50 measure generated by Gold et al. simply because their database is being used. A multistage analysis should be applied to those data. This will reduce errors introduced when data sets are non-linear. The expedited approach of Hoover et al. involves using the Gold et al. database as a tool for identifying studies for analysis, and also using the tabulated data as the basis for analysis in most cases. The exception is for cases where there is high intercurrent mortality among treatment groups. In such a situation an attempt should be made to obtain individual animal data and a time dependent multistage analysis performed.

It is suggested that the same approach be applied here. Selection of a suitable bioassay data set may come from either the Gold et al. database or through literature review but would feed into the same dose response analysis. It would be more expeditious to start with the Gold et al. database and supplement that with additional good studies identified in the literature.

Gold et al. have already applied a series of criteria to the available data; those criteria are not as stringent as the ones laid out in the document. It is noted though that the thorough evaluation of studies with respect to criteria laid out in the document and documentation that such criteria are met will not result in an expedited approach. Also, stringent application of criteria may result in exclusion of valuable datasets, especially if the requirements that the bioassay contain multiple doses (as indicated in the Executive Summary) and peer review are adhered to. Another concern is the strict application of route criteria. Perhaps instead the conceptual approach for cancer dose response analysis should ask -- what is the best approach for characterizing the carcinogenic activity for this chemical? The criteria should be used as a guide for selecting the best studies but should not preclude the use of the dose response data or force a TD50 estimate when they are not met.

Non-cancer dose response evaluation

The cascading scheme for determining non-cancer benchmarks appears reasonable and defensible. Succession of methods used should proceed from those producing the most certain result to those of less certainty. It is not clear that this is necessarily the case in terms of the order of the chemical analog analysis and the TLV/RfC regression. Some further exploration of this issue to evaluate which one results in more certain estimates is needed. Again, too strict application of the criteria for study evaluation may have the unintended result of moving one toward a less certain analysis and result. As for cancer endpoints, the criteria laid out for study selection should be characterized as study selection guidance. Some minimal data requirements for dose response analysis will be needed, but they should be far less stringent than those suggested by

the criteria.

Charge 4: *For constituents lacking dose-response data, limited use of chemical analogs has been proposed. Examples used by EPA include toxicity equivalency factors (TEFs) developed for dioxins and PAHs, as well as the use of a sulfate or salt form of a compound as a surrogate. Do you believe that, given the goal of the expedited methodology, these approaches are scientifically defensible or should be expanded to include additional techniques and constituent groups?*

Dr. Hoover: The use of strong acid salts in place of the parent compound has been widely applied based on the presumption that in either case the same active form is present. Similarly, use of TEFs for dioxins and PAHs is also widely accepted. I agree with the approach to use expanded QSAR on a case by case basis.

Dr. Juberg: They should not be expanded at this point until additional information is available to warrant their use. This approach needs to be very careful with the use of equivalency factors, and perhaps even more careful with the use of sulfate or salt forms of a chemical as a surrogate. This should only be done on a chemical-by-chemical basis.

Dr. Pascoe: The approach described for using chemical analogs where dose-response data are lacking is a conservative approach. On the surface, the limitations that are planned appear to be more restrictive than necessary. However, there are no additional chemical groups to which a QSAR-type approach could be applied in a clearly defensible manner. Data are insufficient to comfortably determine what other chemical groups or moieties could be used in a relationship with toxicity. As such, the approach as planned should be considered sufficient.

Dr. Zeise: As discussed above, the Gold et al. database is a valuable source of compiled dose response data to which a multistage analysis should be applied. In cases of high intercurrent mortality time dependent analyses should be performed if data are available. In application of the approach standardized rules for correction of study length (assuming say risk increases with the third power of age) and adjusting for route of exposure should be employed.

Stringent application of criteria for study selection is not scientifically defensible. The resulting exclusion of studies will push the evaluation toward less certain techniques. For example, the document states that for inhalation benchmarks only inhalation or intratracheal routes of exposure will be used. For the case for which there is only an oral bioassay for cancer and an inhalation number is needed, consideration given to route adjustment. Also stringent application of criteria may result in exclusion of valuable datasets, especially if the requirements that the bioassay contain multiple doses (as indicated in the Executive Summary) and peer review are adhered to. The criteria instead should be used as guidance, with the overall objective of obtaining the most defensible estimate of potency given the data sets and techniques available for any particular case.

Also, the expedited methodology should include a component for expedited cancer hazard identification, also discussed above. Use of information on different organic salts which dissociate to the same active form is appropriate, and the same logic should also apply at the hazard identification phase. Care should be taken in evaluating the quantitative activity of metallic salts; mischaracterizations can lead to significant misestimations of cancer potency. The use of TEFs in the context of the OSW program is encouraged as is the expansion of TEFs beyond dioxins and PAHs to other classes of chemicals. Research to establish basic relationships on which to establish TEFs for other classes of chemicals is encouraged. Further expansion of the expedited approach to include toxicity based schemes for those chemicals for which bioassay data are limited or non-existent is also encouraged. Greater exploration of structure activity relationships for use in expedited cancer potency assessment is also encouraged. Such approaches might be more accurate if integrated with toxicity based approaches. More work in this area is clearly needed.

Charge 5: *Following interim benchmark development by one or more of the expedited methods, the methodology suggests that an expedited review process be conducted. At this point, the expedited review has only been conceptualized. Since this review could be accomplished in a wide variety of venues, please provide recommendations for an expedited review process that you feel would meet reasonable standards for scientific rigor.*

Dr. Hoover: I agree that expediting the review process is critical and any number of options in this regard would be acceptable; I provide a few general recommendations as follows. To simplify the process of review, I recommend that documentation on each chemical be kept minimal with the main focus of the review being the methods (the feasibility of minimizing the documentation will depend in part on the final selected approach). I suggest that the method and results first be internally reviewed within the relevant EPA program office related to the HWIR. A quality control process should be employed to ensure the accuracy of the calculations. The method and results could then be provided to a small panel of outside experts for review and comment. Review and comment should be provided in writing. If required, the experts could meet with a facilitator present to resolve contentious issues. Public comments, again with a focus on the method, could be solicited through a notice and/or a public meeting. In terms of detailed comments from the public on a specific chemical, the main avenue of addressing such comments should be through an in depth review of that chemical by the agency upon request by an interested party. The interested party would have to provide a written justification for the request for an in depth review, including relevant studies not already considered by the agency. The program office could choose to proceed with the review if the request is supported by a reasonable justification (guidelines defining “reasonable justification” would need to be determined). The program office would have some time period over which such a review could take place or could also defer to another program office that has already identified the chemical as being under review or has scheduled the chemical for review. In either case, the interim benchmark would be retained until the in depth review is completed. In this way, the goals of the expedited regulatory process would be met and at the same time stakeholders’ concerns would be addressed.

Dr. Juberg: Process review must include external (to the Agency) scientists. In addition, the review must include some mechanism for insuring that a developed benchmark is achievable within the environment, that it is not below background or natural levels (e.g., metals, naturally occurring PAHs), and that it is not overly conservative so as to decrease the utility of this approach as one that is effective and important in development of human health benchmarks. Finally, the review process should require that once a benchmark is developed, that a review of the database for the specific chemical be scanned or reviewed to insure that the database is consistent and supportive of whatever benchmark is proposed.

Dr. Pascoe: A permanent, and paid, external review panel should provide a faster avenue for review than using internal agency personnel.

Dr. Zeise: Clearly the development of expedited values should not be subjected to the same level of peer review and external scrutiny as consensus IRIS values. The document should provide more definition to the review than given in the report. It is recommended that internal review of values be provided by the National Center for Environmental Assessment NCEA on an expedited basis, and that they also undergo limited external review.

CANCER ISSUES FOR REVIEW

Charge 1: *For constituents for which one or more dose-response carcinogenicity studies exist, the linearized multistage model will be used to estimate the CSF. The uncertainty in the interim benchmark will be characterized by considering the additional uncertainty introduced by the use of the expedited process. The expedited cancer benchmarks will be adjusted by an R factor estimated from a distribution of ratios of cancer benchmarks developed by both an expedited and traditional approach. Do you agree with the proposed use of the R factor to account for the uncertainty inherent in an expedited approach and, if not, please recommend alternatives for consideration by OSW.*

Dr. Hoover: I do not agree with the use of an adjustment factor. First, the expedited approach applied in Hoover et al. is NOT the same as the approach being applied in the current document. The approach in the current document is virtually identical to the traditional approach and if this approach is retained it would be inappropriate to apply a factor to adjust the results (see page 13 of these comments for more discussion on this point). Second, even if the Hoover et al. approach were adopted, the adjustment factor would still be unnecessary. On average, the expedited potencies are more conservative than the conventional potencies not less. There are a few cases where the expedited potency is significantly lower than the conventional; these outliers were explained in Hoover et al. There are also cases where the expedited potency is significantly higher than the conventional one. Applying an uncertainty factor would be unnecessarily conservative. Finally, Hoover et al. found that the differences between the expedited versus conventional

potencies were comparable to the differences between cancer potencies derived by California versus EPA. Thus it applying the "R" factor to adjust expedited potencies would be comparable to applying a factor to traditionally derived potencies as a means of addressing uncertainties related to differing scientific judgment between agencies. Such an adjustment has not been considered necessary.

Dr. Juberg: The R factor is not a commonly known or used entity in the risk assessment arena, particularly in the establishment of environmental health benchmarks. One comment on the use of the LMS model – while this has traditionally and historically been the model of choice for the Agency, there has been a recent shift in scientific consensus among some scientists and cancer biologists to suggest that a threshold exists for some cancers and that the default use of this model (LMS) may be inappropriate in some situations. Much research has been conducted on certain chemicals, such as formaldehyde, chloroform, to name a few, that has advanced greatly our understanding of the carcinogenic potential and process, and for chemicals such as formaldehyde, cancer risks have been dramatically and greatly reduced from previous estimates using the LMS. The salient point is that we must continue to use all of the available scientific knowledge during the establishment of any human health benchmark. If the R factor is going to be used as a default measure to account for uncertainty in this approach, there needs to be a validation of the value derived (using the R factor) against background or natural levels in the environment to insure that this factor is not overly conservative and has resulted in an interim health benchmark that is below ambient levels.

Dr. Pascoe: I think that the use of an uncertainty factor, with the linearized multistage model, based on the ratios of cancer benchmarks developed by both approaches is very appropriate, and adds just enough conservatism to justify the expedited purpose of the benchmarks.

Dr. Zeise: The proposal does not follow the expedited approach of Hoover et al. and the R factor does not represent an appropriate adjustment. Please note that overall the expedited method discussed by Hoover et al. resulted on average in more public health conservative estimates than the traditional approach. This is in part due to standardization of the practice of potency estimation subsequent to the development of some of the potency values used by the Agency. The quantitation and data selection procedures for the expedited approach should mirror as much as possible the current traditional approach. Thus it would be preferable to rely on literature reviews and the Gold et al. database for the identification of the benchmark study and to apply a standard modeling procedure to the set. This would include a standardized approach for adjusting for study length, lack of fit of dose response relationships, dose calculations, and interspecies adjustments. No additional factor is needed.

Charge 2: *The relationship between TD_{50}s and CSFs (and/or URFs) demonstrated through regression analyses is proposed for use in developing cancer benchmarks (i.e., the CSF and/or URF is estimated as the lower 95% confidence interval from the TD_{50}). Do you agree that this is an appropriate technique to derive interim cancer*

toxicity values? If not, do you have recommendations to improve upon this technique or can you suggest other methods that you consider to be more appropriate for this application?

Dr. Hoover (Charge 2 and 3 together): As noted above, if a TD50 is available, then a cancer benchmark study is available and there is no need to use the regression relationship. I do not agree that the TD50/CSF is significantly less resource intensive than applying the LMS model. It is a simple matter to identify appropriate studies from the CPDB and run the default model and, as discussed in the Annex to Hoover et al., this approach produces more reliable potency estimates that are consistent with the traditional approach compared to potency factors derived based on TD50s.

Dr. Juberg: Obviously regression analysis and evaluation of the TD50s and CSFs through this technique are purely statistical methods. There needs to be some biological meaning and toxicological analysis incorporated in any approach. This technique may be appropriate as a first-tier screening method, but again, should be followed with more biologically meaningful in-depth analysis depending on the data and time available.

Dr. Pascoe: An argument that the use of the linearized multistage model will be too difficult or time-consuming to be appropriate to an expedited benchmark development process has not been sufficiently developed in the report to rule it out as an approach. An estimated effort for using the cancer model and using the TD50s should have been better explored in order to provide more justification for the TD50 method. Because the regression for TD50 and CSF values is so strong, it should be retained as a next tier method if the dose-response data are insufficient for use in the multistage model.

Dr. Zeise: The TD50-CSF relationship should not be used. Rather the multistage model should be fit to the dataset used to derive the TD50 values and standard adjustments for study length, lack of fit of dose response relationships, dose calculation, and interspecies differences in body size made.

Charge 3: *The use of the TD_{50}/CSF relationship provides a much quicker, less resource-intensive method of cancer benchmark derivation than using the linearized multistage model (for constituents that have dose-response data available). Do you think that these advantages are sufficient to justify a change in the methodology such that this method is used in lieu of the linearized multistage model for constituents that have quantitative cancer data available?*

Dr. Juberg: Perhaps. What needs to be done is a comparative analysis for several chemicals, evaluating them through both the LMS and TD50/CSF approach. It is important to determine what the interim values are through both approaches and if one model yields consistently higher or lower values than the other. Obviously with only two models, it would be difficult to determine

which one is closer to a true health benchmark, but this could be evaluated by retrospective analysis of the toxicological database for each chemical. However, this reviewer would support an alternative to the LMS, where appropriate, particularly if it removes some of the unwarranted conservatism inherent in a non-threshold model such as the LMS.

Dr. Pascoe: Section 3.1.3 points out that the determination as to whether a chemical is carcinogenic could be time consuming using the multistage model approach. Since the CPDB can be used to identify carcinogenicity, why not use it for that step of the process, then use the cancer dose model to calculate the expedited CSF? And then include the R factor for a measure of conservativeness and uncertainty in the expedited approach.

Dr. Zeise: I disagree entirely with the characterization of the application of the linearized multistage model as resource intensive. Results using readily available software packages can be produced within a matter of minutes. The use of the linearized multistage procedure (including the approach taken for poor data fits) is clearly preferable and more scientifically defensible because non-linearities will be better addressed. Results will be more consistent with the traditional approach.

ADDITIONAL COMMENTS ON CANCER ISSUES

Dr. Hoover: If the original approach is retained, in which only chemicals already identified as carcinogens by EPA, IARC or DHHS are retained for interim cancer benchmark development, I think it is unlikely that a cancer bioassay would NOT be available because such evidence would have been required as part of carcinogen identification. Thus, the contingencies provided for chemicals without adequate cancer bioassays seem unnecessary (the exception of chemicals identified as carcinogens based strictly on structure-activity relationships is discussed in the next paragraph).

Beyond the issue of locating adequate cancer bioassays, I would also predict that for most carcinogens identified by EPA/IARC/DHHS cancer potency factors would in fact be available already. This would be particularly true if Hoover et al. is used as a secondary source of potencies as so indicated in this document. Chemicals identified as carcinogens based on strong structure-activity information (e.g., dioxins, PAHs) are likely to have TEFs and therefore cancer benchmarks already.

I find that the currently proposed method is not sufficiently distinguished from the traditional approach to offer gain in terms of resources and time. I support instead an approach which involves an expedited determination of carcinogenicity where needed and application of the expedited method of Hoover et al. The Hoover et al. approach could be modified slightly by including an updated literature search to check for more current studies that are not available in the CPDB. The addition of the literature search would compromise the expedited aspect to some degree but the benefits of having a more complete database may offset the disadvantage in added

time/resources. The Hoover et al. approach has already been peer-reviewed, both by expert panels in the context of regulatory review and by scientific peers in the context of publication in the literature. An approach was laid out for study selection, which is a critical aspect of expedited potency derivation. Given the goal of carrying out the HWIR process within a time frame of roughly two years (based on information in the charge to reviewers), I recommend against attempting to develop a new "expedited" approach and instead capitalizing on the efforts of the state of California.

NONCANCER ISSUES FOR REVIEW

Charge 1: *For constituents possessing one or more dose-response noncancer studies, the data will be examined to identify the most appropriate NOAEL or LOAEL. To account for the uncertainty in using a limited data set, probability distributions of uncertainty factors (in Baird et al., 1996; Evans and Baird, 1998; Swartout et al., 1998) could be used to define quantitative measures of confidence for the interim noncancer toxicity values. The resulting adjustment factors serve a similar function to the uncertainty factors used in the traditional approach. Do you believe that the use of probability distributions is an appropriate alternative to the traditional uncertainty factors given the limitations of the supporting data set? If not, would you recommend specific changes to how the probability distributions are used or can you suggest an alternative approach that is more technically defensible?*

Dr. Hoover: Generally speaking, a probabilistic approach is designed to be more realistic and less conservative than a default deterministic approach. The use of probabilistic uncertainty factors would not necessarily address the quantification of uncertainty related specifically to the expedited aspect of the approach. The document also points out that using the probabilistic uncertainty factors could have the effect of making the expedited RfDs/RfCs less conservative than the traditionally derived values.

As discussed in the "Highlighted Issues" section, I am in support of the use of probability distributions to represent uncertainty factors along with other improvements to the derivation of toxicity reference values. I also note that under HWIR, probabilistic techniques are applied in the exposure assessment and use of probabilistic techniques on the dose-response side would therefore be appropriate. Uncertainty in dose-response can be more significant than uncertainty in exposure (see Hill and Hoover 1997 for a case study illustrating this point). However, given the goals of an expedited philosophy and the evolving nature of the field, it may not be practical to institute this aspect of the approach at this time. It may be more appropriate to apply the more sophisticated and realistic methods during an in depth review, in which toxicity reference values would be derived to replace the expedited interim values. I suggest that OSW evaluate the advantages of using these methods for deriving the interim benchmarks against the potential disadvantages of the increased

time and resources required, as well as the likely increased difficulty in obtaining consensus on exactly what approach to use in place of the default approach.

Dr. Juberg: Again, there needs to be some comparative analysis done on several different constituents to determine how various interim benchmarks are developed in terms of their end value. The use of probability distributions may serve as a first-tier approach, which could then be followed with the use of traditional UFs which are more constituent specific (e.g., use of various UFs depending on extrapolation from a subchronic to chronic, interspecies sensitivity, etc.). Historically, the traditional approach and use of UFs has appeared to work rather well, although conservatively, and thus unless probability distributions offer distinct advantages, the traditional approach should still be used. It also has enjoyed refinements in its application in recent years and this progress should be included in any application of UFs in this approach.

Dr. Pascoe: Although the Evans & Baird and the Swartout approaches provide a way to deal with uncertainty in the RfD estimates, applying them to an expedited approach with limited data seems inappropriate. Primarily because they result in less conservative RfDs than using a traditional approach. The summary text on page 26 points out the problem with using either approach because of the lack of conservativeness that should be inherent in an expedited approach. However, the Swartout approach still appears to be useful, particularly in the case of limited data, which is probably likely. Why not use the 99 percentile UFs of Table 3-2? These would result in UFs slightly more conservative than traditional approaches, and would retain some of the benefit of an understanding of the confidence in the RfD values.

The discussion on page 26 about the Evans & Baird and the Swartout approaches appropriately points out the advantages of better accounting for the confidence in the RfDs. It may be useful to also point out that there is increasing approval for applying these approaches to the IRIS database to more accurately calculate RfDs. Once that is done, the application of these approaches at a more conservative level to the RfD derivation would be appropriate for an expedited approach. Prior to that happening, the use of the approaches is pre-mature, largely because of the problem of resulting in less conservative RfD values, unless the 99 percentile UFs are used in the Swartout approach, as suggested above.

Dr. Zeise: Although in general the approach of using probability uncertainty factors as laid out by Baird and colleagues may be preferable to the use of the default approach for noncancer endpoints, it does not appear to explicitly address the uncertainties associated with the expedited approach. Potentially the result may be considerably less conservative values than the traditional approach. Although the method shows promise, it is recommended that it undergo further evaluation, particularly within the context of expedited assessment before it is employed by OSW for this purpose. Of particular concern is the lack of data on sensitive human subpopulations (lifetime exposure) for establishing such distributions as well as the lack of understanding of age dependent differences in sensitivity. It is inappropriate to use animal data for derivation of this factor as was apparently done by Baird because of the homogeneity in research animals.

Charge 2: *Lacking the requisite dose-response data, or information on a chemical analog, the next option for estimating a noncancer inhalation benchmark is based on the regression of threshold limit values (TLVs) against RfCs. This regression has been performed and correlation coefficients developed corresponding to specified confidence limits. Do you consider this an appropriate use of these data and can you suggest improvements for this approach?*

Dr. Hoover: I think this is a valid expedited approach. The OSW document should, however, point the potential flaws/uncertainties in this approach. The TLVs have been criticized as being insufficiently health protective (see for example Roach and Rappaport, 1990) and this should be noted. Given the fact that the approach uses a lower 95% confidence bound on the regression, however, I do not consider the potential underconservative nature of the TLVs as a significant issue.

Dr. Juberg: This is purely statistical treatment of evaluating and attempting to correlate values for 8-hr occupational exposure with lifetime permissible inhalation benchmarks. I would be very careful and selective in using this regression technique for this purpose.

Dr. Pascoe*:* I believe that the approach to using TLVs to approximate RfCs is the best one currently available for an expedited approach, and is a popular approach frequently suggested to regulatory agencies. It should be pointed out that the approach is appropriately conservative by using the lower confidence limit, especially given the intent of the TLV to protect workers exposed under different conditions than the public, but that it could result in an RfC up to 3 orders of magnitude different than a value determined by traditional methods.

Dr. Zeise: The approach appears to be reasonable, especially given the use of lower confidence limit for predicting the RfC. Please note that the uncertainty in using this procedure may be less than that for the chemical analog. The magnitude of the uncertainties associated with the two different process should be more fully evaluated before settling on which one is preferable in the derivation hierarchy (e.g., Figure 3-1).

Charge 3: *Representative distributions have been developed for current RfDs and RfCs that can be used to establish interim benchmarks for data-poor (i.e., no appropriate studies available) constituents at a 95% confidence level. Naturally, there are disadvantages to applying a single value to a group of data-poor chemical constituents (e.g., likely to be overly conservative for some constituents). From a technical standpoint, do you agree that this approach achieves the desired goal of the methodology, namely, that the interim benchmark is equal to, or more conservative than, the traditional benchmark at a specific level of confidence? If you disagree, can you recommend alternative approaches (or improvements) to*

address these data gaps? Note: Refinements to the distributions based on chemical class are discussed below.

Dr. Hoover: I support the use of current distributions to estimate an RfD/RfC for constituents without data, and have recently applied a similar (though less conservative) approach in a risk assessment of persistent compounds in Canadian breast milk (Hoover, 1999). I think the use of the 5th percentile is sufficiently conservative for an interim benchmark, and is most likely overconservative (as expected based on the definition of the 5th percentile).

Dr. Juberg: Yes, technically, the approach probably achieves the goal, but I believe the goal needs to be reexamined. The primary goal should be to develop a robust, time-efficient, and scientifically accurate and defensible methodology for evaluating and developing human health benchmarks for exit concentration criteria, among other applications. The goal should not be to maintain conservatism in a particular approach, but do use science and scientific principles to our advantage in applying such knowledge to issues of public health.

Dr. Pascoe: I support the use of the proposed distribution method for the intended purposes of an expedited derivation of a benchmark, especially given a lack of alternatives that offer more confidence than the proposed method. In this regards, the cautionary statement mentioned under General Issues above about the likely use of the interim values is important.

Dr. Zeise (Charge 3 and 4 together): I agree with the use of distributions in cases where data are absent. It is similar to the concept of toxicological insignificance applied to food additives. It would be preferable to augment the approach with chemical class information. The approach laid out in the document for doing this appears reasonable.

Charge 4: *The representative RfD and RfC distributions may be modified and improved by subdividing the universe of chemical constituents into classes based on physical and chemical characteristics (i.e., use of SAR information). The methodology proposes a number of distributions for use in estimating interim benchmarks as a function of chemical class. Using the scheme, the interim benchmark is chosen from the class distribution rather than distribution of all constituents. Do you think that the use of SAR in this scheme is an improvement to the use of distributions given the uncertainty associated with chemical/physical properties and toxicity mechanism of action? Do you agree with the chemical classes developed to refine the distributions? Can you recommend changes to the chemical class scheme that would improve the defensibility of this approach? Can you recommend criteria other than chemical class that might be used to refine the distributions?*

Dr. Hoover: In principle this is a good idea. However, given the well known difficulties in using

available data on a chemical class to predict toxicity for a new member of the class (i.e., an untested chemical), it may be most straightforward to use the full data set. This is particularly true given the goal of the approach, which is to produce an interim benchmark quickly and within the constraints of schedules and budgets. I have particular concerns with the small groupings (i.e., n < 10). The tail of the distribution would not be well defined with such a small data set. I also disagree with the use of a catch-all category for "miscellaneous pesticides". Has an expert in structure-activity reviewed the approach in detail? Have sensitivity analyses been conducted by, for example, using alternative groupings determined by a second expert? One option would be to default to the full data set and deviate from this default only for chemical groupings with large underlying databases and strong evidence of the structure-activity relationship. Another option would be to derive an RfC/RfD based both on the overall data set and the appropriate chemical grouping and select the more conservative value.

Dr. Juberg: For individual chemical classes, the use of SAR is relevant and appropriate. SAR should not be used for all of the constituents collectively, and chemical class independence should be maintained. Finally, I would not place an overemphasis on the use of SAR within this overall approach. It remains a tool for analysis purposes, but should not be used as a stand-alone default method for evaluating constituents and developing health benchmarks.

Dr. Pascoe: The use of the SAR scheme to the distributional analysis of RfCs and RfDs is definitely an improvement to the approach, and removes an initial concern about the large range of values that would otherwise result. No other chemical classes would improve the analysis, mostly because of the low number of chemicals that could possibly be placed in any further breakdown of classes.

ADDITIONAL COMMENTS ON NONCANCER ISSUES

Dr. Hoover: It is well known that the results of traditional toxicity assessment on can vary widely depending on the assessor's scientific judgment and the data he/she had at the time of the assessment. Vermeire et al. (1999) cited a study by Dourson and Lu of noncancer benchmarks (i.e., RfDs published by EPA and ADIs published by WHO) which illustrated this point (38 chemicals had RfDs within a factor of 3 of the ADIs, 20 within a 3 to 30-fold range, 6 within a 30 to 300-fold range, and 1 that had an RfD that was 700-fold different from the ADI). This study emphasizes the importance of making the criteria for selecting, evaluating, and applying uncertainty factors to the relevant studies explicit and transparent, such that multiple reviewers would be less likely to produce widely divergent results. This would be particularly important in the context of an expedited approach because such an approach does not allow for the luxury of full consideration of all related data.

As discussed in Vermeire et al., one method for clarifying the noncancer risk assessment process is to develop default assessment factors that are backed up by extensive scientific discussion. The context for the current assessment would not allow such a resource-intensive

approach, but Vermeire's comments do provide a general framework that could be followed to expedite the review of noncancer studies. It will be crucial to distill the study review process down to a checklist or some such simplified method, in which the reviewer only has to determine particular characteristics of the study in order to proceed through decision points. For example, a reviewer might answer questions related to study duration, availability of NOAEL vs. LOAEL, etc., and use those answers to assign appropriate default uncertainty factors. It may be necessary to assign the initial review of each study to one reviewer, with a series of checks by other individuals, to ensure consistency of the reviews.

I suggest that chemicals to be evaluated for noncarcinogenicity be grouped on the basis of the available databases for the chemicals. An expedited approach will be simple for chemicals with one or two studies, and these chemicals should be reviewed first. For chemicals with extensive databases but no noncancer benchmark, there is likely a complicating issue that has prevented derivation of such a value and this would need to be investigated. Applying an expedited approach to such a database would be more difficult, though certainly still possible.

If a chemical has limited data available, I would advise using that data as the basis for a benchmark and comparing the results to those obtained using a statistical approach. The more conservative value of the two could be selected as the interim benchmark.

When a benchmark is derived using a statistical method (in the absence of sufficient chemical-specific data), physicochemical properties/structure activity relationships should be evaluated to determine whether the chemical may be inferred to be "unusual" (e.g., potential for being unusually toxic). If a chemical is deemed to be potentially unusual, calling into question the applicability of the statistical approach, the interim benchmark can still be adopted but recommendations to carry out additional toxicity studies and an in depth review of the chemical should also be made.

When a conservative, statistical approach is used to derive a benchmark, it would be useful to conduct a "real world" check of the value to determine the applicability of the benchmark (e.g., use the interim benchmark to calculate risks associated with natural background and/or typical exposure levels, if sufficient data are available). If the real world check indicates that the benchmark derived is potentially overconservative, the interim benchmark can still be adopted but recommendations for additional toxicity studies and an in depth review of the chemical should also be made.

PAGE-BY-PAGE COMMENTS

Dr. Hoover: I include specific editorial comments below, as well as general comments on approach where relevant. Depending on which approach(es) is ultimately selected, some of the specific editorial comments may not be relevant to the final document.

p. vii. First bullet under carcinogen class make consistent with later reference and change to reflect final adopted approach.

p. vii After bullets, change sentence to "If none of these criteria is met..."

p. viii 4th bullet change to "..the linearized..." (not "linear")

p. ix Clarify cases where the use of the TD50/CSF regression would apply. As discussed above, based on my reading of the document this would only occur when the cancer benchmark study which underlies the TD50 is considered by the OSW criteria to be "inadequate" for dose-response analysis. Or, the intent may be to replace the use of the LMS model with the TD50/CSF regression. I disagree with either approach, but in any case the document should be clarified and made consistent in this regard.

p. x 3rd bullet, change to be consistent with page viii 2nd bullet, i.e., either use "oral or inhalation" OR "drinking water, food, gavage, or inhalation"

p. x. NOAEL is the "no **observed** adverse effect level"

p. x Clarify what other uncertainties are considered to be uniquely related to the expedited approach. Database deficiencies are the most obvious concern, but this uncertainty is also a potential problem in the traditional approach.

p. xi Rearrange sentence after bullets to clarify meaning - here is a suggested revision, but check for appropriate meaning: "Examples include the use of pyrene as a chemical analog for other noncarcinogenic PAHs and the use of one isomer as a chemical analog for a second isomer (e.g., o-nitrotoluene for p-nitrotoluene)."

p. xii First sentence under "Expedited Review Process": Remove "as well" at the end of the sentence (redundant to "In addition" at the beginning of the sentence).

p. xii Third sentence under "Expedited Review Process" - if there are "major points of contention" this would made application of an expedited process even more difficult than a traditional one potentially.

p. 1 Provide more definitions ("waste concentration limit" "exit criteria"), citations, background on HWIR, etc.

p. 1 Sentence before first set of bullets: "... predicted risk**s**..."

p. 1 3rd bullet, delete parenthetical (not needed/informative here).

p. 1 Clarify tiered use of data sources. According to what I found in the HWIR proposed rule,

OSW consults IRIS, HEAST and then NCEA (provisional EPA benchmarks). Is the use of ATSDR and Cal/EPA proposed here or has this been endorsed/applied previously by EPA? The way this is written it sounds as though the use of these alternative sources is standard practice. If this were the case or even if this is proposed in the current document, this raises an additional issue related to application of Cal/EPA expedited potency factors. Specifically, the level of data/effort required for deriving a Cal/EPA expedited potency is less than what is proposed for the "expedited" approach in this document. That would mean that a less rigorous value (i.e., the Cal/EPA expedited potency) would take precedence over development of an interim benchmark based on the proposed methodology. As noted throughout these comments, I would support the use of currently available expedited potencies with application of the Hoover et al. approach to fill in missing values. If this proposed revision to the approach is adopted, then the logic problem mentioned in this paragraph would not apply.

p. 1 What is the implementation scheme referred to in the last sentence? Describe and provide appropriate citations.

p. 2 Rewrite last sentence. Current sentence is a run-on sentence and does not appear to say what I believe was the intended meaning. Suggested rewrite: "By taking the approach proposed here, the Agency can calculate risks based on interim benchmarks while leaving open the possibility of changing/refining the interim benchmarks as more information becomes available. The interim benchmarks are designed to be conservative, such that once a refined/final benchmark is derived predicted risks would be lower than those based on the interim benchmark." Shorten and refine as needed.

p. 3 An NTP bioassay report is an original study, not a "secondary data source".

p. 4-5 Use of terms such as "critically evaluated", "adequate" - careful definition of what these terms mean is needed.

p. 6 First full sentence suggested rewrite: "Long-term carcinogenesis bioassays in rodents usually include at least 50 animals per sex in each of three treatment groups and in a concurrent control group. The duration of the long-term bioassays is usually 18 to 24 months."

p. 6 Third sentence, second full paragraph - provide citation for statement that less than 15 days is an acceptable duration for developmental studies.

p.6 Sixth sentence, second full paragraph, add word: "Mortality or endpoints relevant to cancer are not selected...as these **endpoints**..."

p.7 Provide the number of constituents rather than saying "Given the number of constituents..." Are these constituents known? If so, include list of them as an Appendix. This would allow reviewers to suggest data sources and approaches for evaluating these chemicals more effectively.

p.7 What are the regulatory goals? Spell them out or at least provide relevant citation. Providing sufficient background is essential for complete and effective review as well as clarity in the document.

p.7 Last sentence, first paragraph add phrase: "...that formally incorporates consideration of the larger uncertainties that will be introduced **by applying an expedited approach rather than a traditional approach**." As well, define better what the "larger uncertainties" are.

p. 8 Second full paragraph first sentence, suggested rewrite - "...the expedited approaches **proposed in this document**..."

p.9 As noted above, Figure 3-1 is not consistent with the figure on page vi. Revise one or the other for consistency.

p. 10 Ensure the listed classifications are consistent with earlier information in the document (i.e., EPA Group A, B or C, OR EPA Group A or B).

p. 10 Which chemical constituents of concern for the HWIR rule have been identified as carcinogens by EPA, IARC or DHHS but do not yet have a cancer benchmark? I would guess that since the identification process is slow, most of the chemical carcinogens would have a benchmark available from one of the listed data sources.

p.10 First sentence under 3.1.1: What exactly is "a shortened version of the traditional process"? How will the traditional approach be shortened? The literature search does not seem to be any different from the traditional approach - how will it be determined which studies will be reviewed and which ignored in order to achieve an "expedited" process?

p. 11 Second sentence, second full paragraph: Though this sentence states that "one or two" studies for each constituent will be collected and analyzed, no details on exactly how this limited database will be selected are provided. The best way to clarify this approach would be to provide the full list of constituents identified as carcinogens and work through a few examples in an Appendix. It may be determined for example that there is only one bioassay for many of the carcinogenic constituents that do not yet have benchmarks from EPA or secondary sources. Then the issue of restricting the database becomes moot.

p. 12 Fourth full paragraph, sentence before bullets: This sentence clarifies the fact that RTI did the analysis on the "R" values based on data provided by Hoover et al. This should be clarified in an earlier description of "R" (i.e., that the underlying data was obtained from Hoover et al., but the analysis was carried out by RTI).

p. 13 First full sentence: The process proposed in this document is NOT essentially the same as Hoover et al. Probably the most significant means employed by Hoover et al. for expediting the CSF estimation was the elimination of the literature review. As noted on page 269 of the article

"Use of information in the [Carcinogenic Potency Database] can facilitate several tasks that consume considerable time in conventional assessments: performing the literature search, collecting all relevant data sets, excluding poor data from the analysis, and identifying the most sensitive species." In the current approach, the requirement for literature review is maintained making it significantly different from Hoover et al. This is a key difference in terms of interpreting the R value. Some of the explanation for the differences between the CSFs derived by Hoover et al. under the expedited approach as compared to the traditional approach lay in the literature review. For example, in one case the regulatory agencies rejected the most sensitive study available based on knowledge gained in the literature review while the expedited potency was based on this study. In other examples, the studies (including epidemiological) used by the regulatory agencies were not available in the CPDB. If a literature review had been performed, these discrepancies would likely have been resolved and the ratios between certain expedited and conventional potencies would have been closer to 1. Thus, as pointed out above, the use of an adjustment factor based on "R" to CSFs derived using the somewhat restricted traditional approach proposed in this document is likely to be unnecessarily conservative.

p. 14 Existing TEFs have already been applied to derive corresponding cancer potencies in numerous circumstances. Similarly the use of strong acid salts and isomers is accepted. I would consider these as standard, accepted methodologies and don't think it's necessary to highlight them as being part of the expedited approach.

p. 16 First sentence, second paragraph: "Factors considered in selecting the TD50...include...elimination of a potentially time-consuming process for determining if the constituent was a likely carcinogen." This statement is not consistent with other parts of the document, as the only time a TD50 would be used to derive a CSF under the proposed approach would be if the constituent had already by identified as a carcinogen under the "time-consuming" review processes of EPA, IARC or DHHS.

p. 16 Third and fourth sentences - Justifying the use of inverse CSF or URF by stating that "a low TD50 indicates a high potency whereas a high CSF or URF indicates high potency" is not necessary for a scientific audience. The fact that the TD50 is a dose, whereas the CSF/URF are in units of risk per unit dose makes apparent the need for the inverse variable.

p. 17 Grammar of last sentence first paragraph: ..."however, advantage**s of the regression approach include that**..."

p. 17 General comment: As noted in the Annex to Hoover et al., a simple formula (properly scaled to adjust for differences in the treatment of experiment duration between EPA and Gold et al.) can be used to derive the CSF from the TD50 rather than applying the regression. The formula is best applied when background incidence is negligible but still produces reasonable estimates even when background incidence is not negligible. As discussed in Hoover et al., however, the most reliable approach to deriving a cancer benchmark is to run the multistage model using the cancer benchmark study underlying the TD50.

p. 21 The difference between the approach proposed to derive interim noncancer benchmarks and the traditional approach is not clear. There are no clear guidelines provided to explain how the database of studies to be reviewed for deriving the interim benchmark would be limited.

p. 24 Accounting for incomplete database - how is it proposed to translate this information into an adjustment factor? This is not provided in Table 3-1 (as it was apparently not addressed in Baird et al.).

p. 25 It is not clear why Baird et al. would advance the stated goal of ensuring that the interim benchmark derived using expedited methods would be conservative relative to the traditional RfD/RfC. Similar comment for Swartout et al.

p. 27 I suggest expanding the citations for petroleum hydrocarbon surrogates [see end of these comments for additional references].

p. 27 Why is this discussion of chemical analogs expanded and different from the one provided under the cancer benchmark section? Why is the acetate salt now included, whereas before only strong acid salts were mentioned? I suggest that the entire discussion of chemical analogs and expanded QSAR be taken out of the sections on cancer and noncancer benchmarks and incorporated into one section on this topic.

p. 30 Third sentence - Since the document is dated June 1999, I am not clear on why the data are current only to August 1998. I suggest updating to current database prior to finalizing the document.

p. 30 Second bullet, delete first parenthetical not needed.

p. 30 Second bullet, second parenthetical - Based on previous discussions in the document, it is not the “true RfD” but rather the “RfD derived using traditional methods.” So suggested rewrite:, “...to be 95 percent confident that the new chemical would have an RfD derived using traditional methods at or above the RfD assigned by this expedited method.”

p. 30 What is the basis for selecting the 5th percentile? Will this be evaluated/adjusted using input from risk managers?

p. 32 Section 3.2.2.5 - An expert on SAR should be consulted. As discussed previously, sensitivity analyses should be run to consider the impact of alternative groupings. I also suggest carrying through a few case studies as examples.

p. 42. Hoover et al. is incorrectly cited. Title should read “Improving the regulation of carcinogens by expediting cancer potency estimation.”

Appendices in general: Run spell check (e.g., p. D-1, D-14, D-15 "halogenated"; E-21, "miscellaneous")

Appendix C: For better ability to review, sort on chemical name rather than RfD/RfC.

Dr. Juberg:
Page v – Within the first paragraph, the text should state that the proposed approach is purposely conservative.

Page v – 2nd paragraph – Text might include a sentence that removes any question as to why human health benchmarks typically do not consider or establish values for the dermal route of exposure. To a lay reader, they may question why this route is not included.

Page vii – "In the absence of a constituent……..since it has not been sufficiently established that the constituent is a carcinogen." Good point here and this reviewer commends the authors for restraining any development of a cancer benchmark when there is absolutely no evidence that a constituent has carcinogenic properties or propensity.

Page viii – "The cancer benchmark study must include two or more doses, with the results showing a statistically significant dose-response relationship." What type of dose-response relationship? One that considers only malignant tumors? Combined malignant and benign tumors? Please specify.

Page viii – "If several dose-response carcinogenicity studies exist, the study showing the most sensitive dose response will be selected." At all costs and despite perhaps, other more meaningful data and information? This seems to place greater emphasis on insuring the conservative nature of the proposed approach at the expense of perhaps valuable and meaningful knowledge or information that might enhance the accuracy and defensibility of a derived value.

Page viii – "The linear multistage model will be used to estimate the CSF." This reviewer simply does not agree that the LMS should continue to be used without question, simply because it has been the Agency's choice historically. There is increasing evidence (and consensus development as noted by a hand vote at the 1998 Society of Toxicology Annual Meeting) that a threshold approach may be more appropriate for many "suspected carcinogens" and that the use of the LMS to evaluate and establish CSFs should be implemented on a case-by-case basis. Be careful with the continued use of a default approach that is decreasing in its utility and favor amongst cancer biologists and toxicologists.

Page viii – 1st paragraph – The whole topic and discussion on the use of the R factor appears non-chemical specific. In other words, again, this approach emphasizes the statistical treatment of constituents and places less importance on the biology of a constituent, the true measure by how and what mechanisms it may impart carcinogenicity. There needs to be some recognition or provisions that permit chemical-specific information to be considered in deriving interim health benchmarks. In addition, there needs to be additional discussion and explanation in this paragraph as to how the R

factor will be applied.

Page ix – "Expanded QSAR" – Given the already tentative reliability of some of the proposed approaches for establishing interim benchmarks, this reviewer sees little utility or need to incorporate "expanded QSAR" into this approach. This exercise seems too far removed from other factors (e.g., mechanism, biological meaning, genotoxicity) that have more utility in predicting a constituent's ultimate toxicological properties and effects. Minimize the discussion on this topic. Should only be used in extreme cases.

Page x – 1st paragraph – Do the authors mean an "interim RfD or RfC" will be developed? This is an expedited approach and this reviewer assumes that an interim benchmark will be developed and not a traditional RfD or RfC which would require much more time and peer review. Please clarify.

Page x – Should the benchmark study only need to include two doses? Why not require three?

Page x – "The results should demonstrate statistical significance….." The results should actually demonstrate effects that are toxicologically and biologically meaningful. Of course, statistical significance is important, but any effect resulting from treatment must be important from a toxicological perspective. Additionally, it is very important that there be some degree of concordance between the effect and results of clinical chemistry and pathology evaluation. These additional observations and data are often confirmatory that an observed change or effect has toxicological significance.

Page x – "For constituents……for the most sensitive or most relevant species." Place greater emphasis on the most relevant species. There is enough conservatism built into this process and approach that the most sensitive species does not necessarily need to be considered as the best choice. Additionally, I would place significant emphasis on selecting a study and/or animal model that is the most relevant TO HUMANS. This may be implicit in this discussion, but should be emphasized.

Page x – 2nd full paragraph – "However, use of an expedited method adds to the uncertainty." Not necessarily in all cases. This sentence is too declarative and cannot stand alone without additional support. Please revise.

Page xi – "Expanded QSAR" Again, this is a stretch and not very useful for this discussion or approach. Minimize or delete.

Page xi – Several comments on the preferred statistical approaches:

"Derivation of an interim RfC from a TLV." Remember that this involves extrapolation of an 8-hr benchmark to a lifetime chronic exposure (e.g., RfC).

"Derivation of …….from the complete distribution….." Again, this is non-chemical specific. Need to confine to chemical class.

Page xii – "In other words, chemicals with interim……." These last three sentences in this paragraph are somewhat confusing. Can the authors please clarify the meaning of this discussion.
Page 1 – "Human health risk analyses performed by OSW require that at least one health benchmark be available….." What if the one that is available is an oral benchmark and inhalation is the only route of concern? Or vice versa? There needs to be some discussion on how to treat this scenario, since it likely will be quite common.

Page 1 – 2nd full paragraph – There should be provisions for also utilizing non-Agency peer-reviewed literature and/or data, particularly if scientific consensus exists for a particular mechanism, effect, etc. The field of risk assessment and safety evaluation has been hampered somewhat by the lack of consideration and use of non-Agency conducted studies and research. Often, these information sources are in fact, of higher quality, since they incorporate GLP, must be conducted according to current protocols and standards, as well as withstand peer review. In contrast, many Agency-produced documents and data sources are not subjected to external peer review and as such, their utility is limited for an approach such as this. This point cannot be stressed enough that a weight-of-scientific evidence approach and incorporation of all relevant data be implemented.

Page 1 – 3rd paragraph – What is meant by "secondary literature"?

Page 1, last bullet – "Develop an implementation scheme that is not directly based on benchmarks." What is the scheme then based on?

Page 2, 1st bullet – Good and appropriate point – "…include additional constituents as EPA-approved benchmarks become available."

Page 2 – "….will require greater use of extrapolation between constituents and species and routes of exposure…." There should be minimal, if any, extrapolation between routes of exposure, given the known differences in metabolism, portal of entry effects, mechanisms, that occur via different exposure routes. There should be a limit as to when, how, and by what means route-to-route extrapolation is conducted.

Page 2 – "A guiding principle adopted…..should be conservative…." NO. This is not a principle, but rather a policy and this should be stated. A principle is a basic truth, law, or assumption. A policy is a prescribed or selected option. Please clarify as science policy needs to be distinguished from fact based on scientific evidence.

Page 2 – "these interim health benchmarks may change as more information…." This is not likely to happen as we have witnessed through the IRIS program. It is extremely difficult, to the chagrin of the scientific community, to change or alter health benchmarks once they are placed in the public domain. I would not lead the readers of this document to the hope that interim benchmarks are likely to be readily or easily changed, unless the authors can provide some justification and proposed mechanism for such.

Page 3 – Again, what is meant by "secondary sources"? Are these peer-reviewed studies? Agency data? CBI data, case reports? There needs to be some standard by which a study or data are evaluated and used. This approach cannot use just any data that exists.

Page 4 – 1st full paragraph – "A literature search….." Good to identify these various resources and again, you should note that non-Agency research and data should be considered as valid sources of information if they have been peer-reviewed and are considered high quality data and studies.

Page 4 – Good notation of the criteria required for studies that are to be considered for use in risk assessment. However, a requirement should also be added that specifies that a relevant route of exposure (to the human situation) be used if available.

Page 4 – "Exhibit a dose-response relationship" This should not be an explicit requirement, since a well-conducted and high quality study may not, in fact may infrequently, demonstrate a dose-response; yet, these data may still be very useful and important from a risk perspective. Be careful about requiring a dose-response demonstration at the expense of neglecting a better study that is useful, yet did not demonstrate such. There still is merit in such a study for safety evaluation and risk assessment purposes.

Page 5 – Add a bullet that notes the utility and benefit of understanding toxicological mechanism of action.

Page 5 – 1st full paragraph – Excellent point about the mechanisms in various animals that we have learned about and which are not appropriate to humans. These should not be used as a basis for a human health risk assessment. This is the type of current information that needs to emphasized throughout this document. Other examples include consideration of the 1996 revised cancer RAGS and the use of PB/PK information when available.

Page 6 – "benchmark study must be conducted on humans…." Rarely are human studies prospectively conducted and I would revise/rephrase this sentence.

Page 6 – "If several cancer benchmark studies meet the above criteria, the study indicating the highest cancer potency is selected." Again, at all costs and regardless of the utility and perhaps more toxicological relevance of other studies? Does this position hold, regardless of the quality of the study? Again, this statement suggests that greater emphasis is placed on policy and conservatism rather than on scientific merit, rationale, and meaning. I don't think the Agency wants to convey the message that this approach is simply a statistical exercise and is placing greater meaning and weight on those studies that yield the most conservative data with which to derive an interim benchmark. This reviewer cannot support this statement from a risk assessment perspective.

Page 6 – "A noncancer benchmark study should be of chronic……" Actually, an appropriate study length will depend on the type of study conducted and for what endpoint of concern. A minor point, but a benchmark study may not necessarily be of chronic or subchronic duration, but still be used if

it evaluated the toxicological endpoint of concern. This is well illustrated in the subsequent discussion about the use of developmental toxicity studies that are not of chronic duration. Please reflect this good example in the beginning sentence as well.

Page 7 – 2nd paragraph – "The traditional….. the assumption that noncancer effects have a threshold whereas cancer effects do not." See comment referring to page viii above. Again, this view is changing as there is a trend developing among many scientists that cancer effects may be threshold-related as well.

Page 7 – 2nd paragraph – "For cancer effects, no exposure level is considered risk free." Again, same comment as above. There needs to be some discussion that brings to light the changing picture with regard to cancer effects and the fact that a no-threshold assumption may not be correct. More importantly, while the 1996 EPA cancer RAGS are cited, there should be some attempt to weave these into the discussion and into the interim approach proposed.

Page 7 – 1st bullet – This is the first mention of a constituent being absorbed by the dermal layer. Do you want to discuss this route of exposure when it has not been considered previously?

Page 10 – 1st paragraph – It is difficult to discern why in 1999, the linearized multistage model is once again selected as the basis for estimating and developing CSFs and URFs. This seems counterproductive, even for an interim screening approach, to the work that has been done in the area of cancer biology in the past decade and before.

Page 11, 2nd full paragraph – "….some subset of those studies (one or two…..)" Do these "studies" refer to cancer bioassays? Please clarify.

Page 11, 2nd full paragraph – This entire paragraph appears very statistically-oriented and almost seems to abandon the science. For the subject matter at hand (i.e., development of interim human health benchmarks), this discussion is too abstract and makes the reader wonder whether statisticians were the primary authors of this section. Please make a case for why there needs to be so much attention and emphasis on statistical analysis and theory.

Page 12, 3rd full paragraph – Same comment as above. Far too statistical in nature. Mathematics and statistics do not govern carcinogenesis or the propensity for a chemical to act as a carcinogen. Minimize this particular discussion.

Page 12, number 2. – The R value of 0.15 seems very conservative. Can an example be shown that demonstrates how this R value is applied and how it alters the final CSF? Please provide an example to show by how much it reduces the final interim benchmark. By two-fold, ten-fold, or greater?

Page 12 – There is some concern about using an R value that has been derived from 70 different constituents and whether there is enough chemical specificity embedded within this value so as to have any meaning when it is applied to specific interim benchmarks. In other words, the R value appears

to be an averaged value based on many different constituents. Perhaps this is the intent and if so, there should be some mention that it is a generic (non chemical specific) value.

Page 13, 1st paragraph – "…will result in human health benchmarks that are overprotective…." " The degree of overprotection is not known, using these methods." This reviewer is concerned that there may be so much protection, or conservatism associated with this approach, that exit levels may be well below background concentrations, and thus meaningless from a risk assessment perspective. If an interim methodology sets a level that is below natural or background levels, then what is the utility of this approach in terms of its true measure as a health protective approach? Again, where has the biological meaning and relevance gone? There is again too much emphasis on statistical analysis in this approach.

Page 13, Boxed area – When a sulfate or hydrochloride form of a constituent is used as a surrogate, it certainly should be chemical specific. There are many such forms that are much different in toxicity than the parent compound.

Page 15, 3rd paragraph – Insert "a" into "….may be performed to produce a provisional…."

Page 15, 3rd line from bottom – Hyphenate "chemical's"

Page 16 – Small detail, but insert "CPDB" after the title is first introduced, as later on it is used standing alone, and the reader has to again determine what it is an acronym for.

Page 17, Section 3.1.3.2 – This brief discussion serves to point out well, how much statistical analysis and emphasis is embedded within the approach. There is simply not enough discussion pertaining to the biology and toxicological properties of various constituents and how these influence a chemical's carcinogenic potential.

Page 21, 2nd bullet – Should this be a "statistically sufficient" number of test subjects?

Page 21, 1st bullet – Unlikely to have a chronic or subchronic study in humans.

Page 21, 3rd paragraph – Excellent point about selecting a NOAEL or LOAEL in the most relevant species.

Page 21, last paragraph – This discussion is all standard knowledge and it is probably not needed here.

Page 22, paragraph beginning with "Uncertainty factors of 1,3……" Again, this is standard and common knowledge and probably not necessary within this document.

Page 22 – "…there is no way of knowing how much confidence should be placed in the claim that any true threshold for effect in the most sensitive human subpopulation will be below the RfD or RfC

value." It should be noted that thresholds may be above the RfD or RfC and there is the possibility that they are far above these health benchmarks.

Page 23, last line – "….more heterogeneous…." In what respect? With respect to genetics, metabolism, hair color, etc. Please specify.

Page 24 – "Pseudo-GSDs, that are the square root of the 84th percentile divided by the 16th percentile, were provided by the study authors because the actual distribution is not lognormal but is characterized non-parametrically." This reviewer has serious doubts as to whether many scientists or risk assessors will have any idea as to what this sentence means. This is a prime example where the approach and documentation contained with this Draft, is too statistical in nature. What does this sentence have to do with toxicity, biological effects, biological meaning, etc.

Page 25 – "Application of the adjustment factors presented by Baird et al. (1996) however, would necessarily produce an interim noncancer benchmark that is less conservative than the standard approach, violating a primary goal of the expedited approach." Yet, what if application of the factors proposed by Baird are more meaningful or accurate from a scientific perspective, ones that yield a value that is closer to the true mark. The policy of upholding conservatism in this approach could not be spelled out more clearly than in this paragraph and this makes the reader question whether the OSW is more interested in identifying and developing accurate health benchmarks or in insuring that any developed benchmark is more conservative than would be traditionally derived.

Page 26 – "…would result in interim noncancer benchmarks less conservative than the standard approach, which would be inconsistent with a primary goal of the expedited approach." Who cares whether or not the approach is less conservative, what we should be interested in is biological meaning, scientific accuracy, and in developing a benchmark that is grounded in science and that approximates the best estimate of a true health benchmark. The policy discussion should be placed elsewhere. The authors appear to be discussing various approaches, but then discounting their value and utility because they violate the "conservative nature" that is desired within this approach.

Page 26, 2nd to last line – "select a benchmark from the distribution of existing benchmarks" Benchmarks for what effects and for what chemicals?

Page 27 – "In general, the toxicities of salt, sulfate, and acetate compounds are not expected to differ significantly from their parent compounds." Be very careful with this statement. There are many such compounds whose toxicity varies significantly from the parent compounds and this is an over generalization.

Page 30 – "to an RfD of 0.00008 mg/kg/day)." This is a relatively low RfD and one wonders whether through statistical application of the log function and the resulting low value if there is any biological meaning in this RfD.

Page 30 "The advantage of this approach is that it is simple, quick, and requires no chemical-specific

information." This is what is worrisome in that there is absolutely no consideration of the biology, chemistry, pharmacokinetics, toxicology, or any other parameter with more relevance that simply statistical manipulation of RfD or RfC distributions.

Page 30 "The interim exit concentrations, however, are likely to be significantly lower than those that would be developed under a traditional approach, and the degree to which they are lower cannot be determined." This is unacceptable as an explanation and the authors must anticipate or forecast what will develop if exit concentrations or interim benchmark values are established that are lower than background, natural, or even detectable concentrations. This serves no purpose and if such is the case, then an interim approach should be abandoned.

Page 35, last paragraph – Better to use ranges of health benchmarks within chemical classes than to use the entire distributions.

Page 38 "Expedited Review" – Obviously, while the authors admittedly are desiring input on this section and perhaps that is why there is a general lack of critical information on the review process, there needs to be some further development of the specifics associated with this part of the process. It is too brief, yet one of the most critical steps in health benchmark development. What is the process that will be proposed? Will external (from the Agency) reviewers be used?

Page 38 – "An expedited review may involve a limited number of reviewers,…. Relaxing the need for consensus, and allowing for greater uncertainty so long as the interim health benchmarks are likely to be lower than those that….." Is this what we want out of such an approach, for their to be no need for consensus? For there to be greater uncertainty, so long as we maintain the conservatism of the developed values? If there are a limited number of reviewers, there had better be a diversity of disciplines represented, particularly risk assessors, toxicologists, and those versed in safety assessment.

Page 38 – "Use of the expedited approach to develop interim cancer and noncancer benchmarks will result in values more conservative than would use of a standard approach." Yet, in actuality, any approach could be used as long as one simply increased the uncertainty factor sufficiently. What makes this approach unique or different from the standard approach that would increase the UF in developing health benchmarks?

Page 38 – "It should be noted that interim health benchmarks may change as more information becomes available……" The Agency must have in place some mechanism or opportunity for external scientists to provide additional information and data that may be useful in changing an interim value. What are the provisions if an external scientist wanted to submit an alternative value based on sound and defensible data?

Page 39 – Change "in this report should be applied" to "could be applied"

Page 39 – Same change in two places "following methodology could be used" and "the following methodology could be used"

Page 40 – Top two lines – This is redundant. Authors have already detailed these points. Omit.

Page 40 – "relaxing the need for consensus" This is not necessarily good.

Page 40 – 1st paragraph – All the same points as previously mentioned: Who will the reviewers be, is it necessarily good to have greater uncertainty, etc.

Page 40 – Last paragraph – This is a poor finale or summary of the entire document and the authors should attempt to bring the focus together at this point. Why end the entire document with a sentence on methods available to extrapolate RfCs to RfDs and from inhalation to ingestion cancer benchmarks?

Appendix A-1 – Is the TD50 geometric mean a mean of both rat and mouse TD50s? In other words, have TD50s from different rodent species been averaged and if so, do the authors consider this a scientifically defensible approach?

Dr. Pascoe:
Section 3.2.1. This section is a little difficult to follow at first. It would help greatly to have two subheadings: "Traditional Process" at the start of the first paragraph, and "Expedited Approach" at the start of the paragraph immediately following the equation for the RfD on page 22.

It is not clearly stated, but appears that the expedited approach would follow the traditional approach described in the first set of bullets under Section 3.2.1, page 21. This needs to be stated at the beginning of the "Expedited Approach" section on page 22, followed by the discussion of the expedited approach to the uncertainty factors.

Page 24, Accounting for Incomplete Database – The last sentence of the first paragraph is a little confusing. It would help to rearrange as "Evans and Baird examined…and calculated how much lower the missing NOAEL might be for each…"

Page 24 - The equation is unclear as to what the "Available minimum NOAEL" refers to. Is this a ratio of the NOAEL from each study to the minimum NOAEL in the database?

Page 24 – A tie-in between the data in the Evans and Baird paper on accounting for a missing NOAEL and the issue of accounting for an incomplete database would be helpful at the end of the Evans and Baird discussion.

Section 3.2.2.1, second paragraph, last sentence – The use of pyrene as a surrogate is repeated at the beginning and end of the sentence.

Section 3.2.2.2 – It is not clear how adjustments to the benchmark would be made to account for differences in pharmacokinetics, etc. Some citations to literature that discusses this would be useful here.

Section 3.2.2.3, first paragraph, last sentence – Suggest adding to the end "…extended period, *and often the values are based on the same toxicological or epidemiological studies*."

Section 3.2.2.3, second paragraph – The logic of the sentences and bullets is unclear. Suggest dropping the bullets and change the text to "…are provided in Appendix B. *Based on the regression relationship, a procedure for developing an RfC from a TLV is to* determine the TLV for the constituent of interest and select…"

Dr. Zeise:

PROPOSED REVISED APPROACHES

Dr. Hoover: In consideration of my foregoing comments, I recommend the following general frameworks for deriving interim cancer and noncancer benchmarks. Explicit details will need to be worked out and will depend in part on exactly which chemicals need to be addressed under the proposed HWIR. Elements of the review process are not discussed further (see page 5 for recommendations).

Interim Cancer Benchmarks

For chemicals identified as carcinogens by EPA, IARC and DHHS and chemicals identified as carcinogens using an expedited determination, apply the approach of Hoover et al. to derive interim cancer benchmarks. The approach involves:

- Use of cancer dose-response data that has been evaluated and extracted from the literature and summarized in the CPDB by Gold et al.;

- Selection of appropriate study from the CPDB based on explicit criteria (adopt from Hoover et al. or modify as appropriate to current goals); and

- Application of the LMS model to the selected study to derive an interim cancer benchmark.

As an optional addition to this approach, use a current literature search strictly to identify newer studies not included in the CPDB. Studies from the literature could be evaluated using the CPDB approach in order to maintain consistency and selected using the study selection criteria. Strong acid salts could be used in place of parent compounds (as well as other salts provided that the active form is the same). TEFs could be applied as appropriate.

Interim Noncancer Benchmarks

A suggested framework for the derivation of interim noncancer benchmarks is as follows:

- Conduct comprehensive literature search on chemicals under consideration for benchmark development. Issue public notices to elicit additional data.

- Establish explicit screening criteria in the form of checklists for reviewers.

- Sort chemicals based on extent of databases. E.g., establish groups of chemicals with no adequate studies, groups with 1-2 adequate studies and so forth.

- Note chemicals that were found to have recent, adequate cancer bioassays that have not been reviewed by EPA, IARC or DHHS under a carcinogen identification process. Use current EPA carcinogen identification guidelines to screen those studies and determine whether an expedited determination of carcinogenicity is warranted. If so, consider this chemical for development of interim cancer benchmark following approach described above.

- Note chemicals that are classified as EPA Group "C", IARC Group "3" and other comparable categories.

- Beginning with chemicals having small databases, evaluate studies using checklist approach. Based on the checklists prepared by reviewers assign appropriate uncertainty factors. Include uncertainty factor to address limited carcinogenic concerns as appropriate.

- For chemicals with extensive databases [to be defined], call on expert reviewers familiar with these chemicals to determine the reason for no benchmark being available and the practicality of applying the expedited approach to such an extensive database. Use criteria established for selecting adequate studies and checklist approach to defining uncertainty factors if considered appropriate (e.g., a chemical may have extensive pharmacokinetic data indicating that the default expedited approach would not be appropriate).

- For chemicals with limited studies and none considered adequate based on screening criteria, derive interim benchmarks in two ways. First, use the limited studies and the default expedited approach to derive a benchmark. Second, use a statistical approach (e.g., select a value based on the full distribution of RfDs/RfCs) to derive a benchmark. Compare the two and select the more conservative benchmark. This approach will encourage the conduct of additional toxicity studies in order to fill the data gaps and reduce the conservatism of the interim benchmark.

- When applying a statistical approach in the absence of any chemical-specific studies, evaluate the properties of the chemical to determine whether unusual toxicity might be expected. If natural background data or typical exposure levels of the chemical are available, use the statistically derived interim benchmark in combination with the exposure data to estimate risks. Based on either of these checks, a recommendation to conduct an in depth review and/or initiate toxicity studies could be made. The interim benchmark derived based on the statistical approach would still be applied until a more detailed evaluation could be conducted.

REFERENCES

Dr. Hoover: Some of the references below were cited in my comments (I have not listed references already cited in the main document such as Hoover et al., Baird et al., Swartout et al.). I also include additional potentially relevant sources published in 1999 that I identified using a citation search (based on references cited in the document). References for the use of surrogates in assessing petroleum hydrocarbons are also provided.

Brand, K.P., Rhomberg, L. and Evans, J.S. 1999. Estimating noncancer uncertainty factors: Are ratios of NOAELs informative? Risk Anal 19(2):295-308.

Carlson-Lynch, H, Price, P.S., Swartout J.C., et al. 1999. Application of quantitative information on the uncertainty in the RfD to noncarcinogenic risk assessments. HERA 5(3):527-546.

Hattis, D., Banati, P., Goble,R. and Burmaster, D.E. 1999. Human interindividual variability in parameters related to health risks. Risk Anal 19(4):711-726.

Hill, R. and Hoover, S. 1997. Importance of dose-response model form in probabilistic risk assessment: A case study of health effects from methylmercury in fish. HERA 3(3):465-481.

Hoover, S. 1999. Exposure to persistent organochlorines in Canadian breast milk: A probabilistic assessment. Risk Anal 19(4):527-545.

Hutcheson, M.S., Pedersen, D., Anastas N.D. et al. 1996. Beyond TPH: health-based evaluation of petroleum hydrocarbon exposures. Regul Toxicol Pharmacol 24(1):85-101.

Roach, S.A. and Rappaport, S.M. 1990. But they are not thresholds - A critical analysis of the documentation of Threshold Limit Values. Am J Ind Med 17(6):727-753.

Roberts, S.M. 1999. Practical issues in the use of probabilistic risk assessment. HERA 5(4):729-736.

Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG). 1997. Selection of Representative TPH Fractions Based on Fate and Transport Considerations. TPHCWG Working Group Series, Volume 3. Prepared for Association of American Railroads, United States Air Force, and the TPHCWG. Amherst Scientific Publishing. Amherst, MA.

Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG). 1997. Development of Fraction Specific Reference Doses (RfDs) and Reference Concentrations (RfCs) for Total Petroleum Hydrocarbons. TPHCWG Working Group Series, Volume 4. Prepared for Association of American Railroads, United States Air Force, and the TPHCWG. Amherst Scientific Publishing. Amherst, MA.

Vermeire, T., Stevenson, H., Pieters M.N. et al. 1999. Assessment factors for human health risk assessment: A discussion paper. Crit Rev Toxicol 29(5):439-490.

ATTACHMENT A

Comments on"*Conceptual Approach to Establish Interim Human Health Benchmarks*"

Sara Hoover
Golden Associates

COMMENTS ON:

"CONCEPTUAL APPROACH TO ESTABLISH INTERIM HUMAN HEALTH BENCHMARKS" PEER REVIEW DRAFT

Document Prepared by: Research Triangle Institute

Submitted to: U.S. Environmental Protection Agency, Office of Solid Waste

Comments Prepared by: Sara Hoover
For: Eastern Research Group Inc.
Agreement No.: 0120.00.017.002-15
Contract No.: 68-W-99-001

INITIAL COMMENTS

The overall goal of the methodology is clear and provides the underpinning for decisions made throughout the document. I am in support of the use of expedited approaches, which have a two-fold benefit as has been pointed out by other authors: 1. Allow for the regulation of previously unregulated compounds; 2. Encourage the testing necessary to develop a less conservative benchmark. The application of the expedited philosophy should be strengthened and made more consistent, however; detailed comments in this regard are provided in later sections.

ABBREVIATIONS

Abbreviations are consistent with those used in the charge to reviewers and the document.

HIGHLIGHTED ISSUES

The main issues identified in the review are highlighted below:

A more detailed summary of the context for this document (i.e., proposed HWIR) along with appropriate citations should be included in the Executive Summary and Introduction. Some of the background information is already available in the charge that was provided to reviewers and can easily be included in the document. Appending the entire list of the constituents known to be of concern under the HWIR would also be useful to reviewers and users of the document.

The approach to deriving interim cancer benchmarks should be revised and simplified to achieve the goals of an expedited method (see specific comments in later sections and a

summary of a proposed revision beginning on page 16).

The document does not provide enough detail on the primary method for expediting the derivation of noncancer benchmarks. The approach to limiting the data set under consideration must be specified and/or alternative options considered (see specific comments in later sections and a summary of a proposed revision beginning on page 16).

I strongly support the use of more sophisticated techniques to derive noncancer benchmarks, including, for example, the use of benchmark dose modeling, consideration of uncertainty in dose-response model and use of probabilistic uncertainty factors. However, even though the science of evaluating uncertainty in toxicity benchmarks has progressed considerably, these techniques have not been widely applied in a regulatory context (Roberts, 1999) and therefore may not be compatible with the goals of the expedited philosophy. The field of probabilistic uncertainty factors in particular is still evolving.

When using statistical methods to derive RfCs/RfDs in the absence of specific data on a chemical, the chemical should be evaluated for potentially unusual properties that could affect the toxicity. It would also be useful to conduct a "real world" check on the toxicity value derived using these methods (see page 10 for more details).

The document would benefit greatly from the extensive use of illustrative examples.

RESPONSE TO GENERAL QUESTIONS

1. Do you believe that the major sources cited (e.g., Hoover et al., 1995) in this report have gained sufficient acceptance in the risk assessment community for application in a regulatory context?

After detailed review of the document it is apparent that the proposed approach to deriving interim cancer benchmarks relies only in a minor way on Hoover et al. Thus regulatory acceptance of the Hoover et al. approach does not seem particularly relevant to the currently proposed approach (comments on revising the approach are provided later). That being said, in answer to this question Hoover et al. was used as the basis for a lower tier of cancer potency factors which have been promulgated as regulatory values under Proposition 65. Thus the values are certainly accepted in that regulatory context. I conducted a search in the scientific literature for papers that have cited Hoover et al. but located only Cranor (1995). I also looked for references to expedited approaches on the Internet but did not locate any. I am not aware of other state or federal agencies that have applied the Hoover et al. approach or the cancer potency factors reported therein in a regulatory context.

With regard to sources such as Swartout et al. and Baird et al., for example, the field of applying probabilistic techniques to establishing benchmarks is evolving and new papers are being published on an ongoing basis. I conducted a citation search on Swartout et al. 1998 and Baird et al. 1996 and found a number of more recent publications (listed in the reference section at the end of my comments). To my knowledge, the use of probabilistic uncertainty factors has not been accepted in a regulatory context as yet.

In terms of sources regarding TEFs, these are well accepted and have been applied in numerous regulatory circumstances.

As a final note, I do not think that lack of precedent should necessarily influence what is done for the purposes of HWIR. The expedited approach applied by Hoover et al. under Proposition 65 had no precedent but was accepted in that regulatory context. The expedited approach selected by OSW should meet the regulatory requirements and goals of the HWIR.

2. Are there alternative methods not presented in this report that you would recommend for use in deriving interim health benchmarks?

The document discusses the major options. I would expand the reference list to include more extensive citation of the literature on petroleum hydrocarbons. The surrogate approach is widely applied in that field. The use of surrogates in that field is also a recognition of the value of an expedited approach, as the need for doing something rather than nothing in the face of large data gaps for petroleum hydrocarbons motivates the use of surrogates. With regard to other possible methods in the literature, I would suggest conducting a current literature search to check for suitable references as the field of risk assessment is a rapidly evolving one.

3. Do you support the logic illustrated in Figure 3-1 hierarchy? If not what changes would you recommend to strengthen the hierarchy?

First, the diagram on page vi illustrates a different hierarchy than Figure 3-1. Most of the descriptions of the process provided in the document are consistent with the figure on page vi (the one exception is discussed below). For example, the document indicates that for chemicals identified as carcinogens, the first step would be to identify whether a cancer benchmark study is available, and if not, then determine if the chemical analog criteria are met. Figure vi is consistent with this. Figure 3-1 is not. Figure 3-1 appears to show that chemical analogs would only be considered if the TD50/CSF regression cannot be done. The meaning of the double arrow in Figure 3-1 connecting the "Cancer Benchmark Study" box with the "TD50/CSF Regression" box is also unclear, though I assume it means these two options are considered to be at an equivalent level. Based on the document, however, it would appear that the first step after identification of a carcinogens should be to evaluate whether a cancer benchmark study is available, if not then consider chemical analogs and if no chemical analog is appropriate then consider use of TD50/CSF regression. On page 16 in the first sentence of Section 3.1.3, however, the process appears to be described differently: "For some constituents that do not have a **cancer benchmark** [as opposed to a cancer benchmark **study** which is the phrasing in the rest of the document] there may be dose-response data that have been used to develop an alternative carcinogenic potency estimate... This statement would appear to be consistent with the logic in Figure 3-1. Once the method is revised and finalized, I suggest preparing one figure that reflects the appropriate hierarchy.

Second, the diagram does not explain some of the key decision points. For example, in the portions of the document describing interim cancer and noncancer benchmark development, the concept of evaluating/analyzing only a "subset" of studies is mentioned. There should be a diagram showing what logic will be used in identifying the subset of studies, if this aspect

of the proposed approach is retained.

Third, I have concerns about the first step in the hierarchy, in which the cancer designation available from US EPA, IARC or DHHS is used to identify carcinogens. Given the lengthy review time required to make a cancer designation by any of the three named agencies, this first step in part defeats the expedited philosophy taken in the current document. For example, suppose an NTP bioassay is available, in which clear evidence of carcinogenicity in animals is supported by statistically significant increases in tumors in two species. If the NTP bioassay is relatively new, the EPA, IARC and DHHS designations are unlikely to be available, yet the likelihood that this chemical would be considered a potential human carcinogen once the reviews have been completed by these agencies is high. I suggest providing for some mechanism to make an expedited determination of carcinogenicity, rather than setting a noncancer benchmark in a clearly inappropriate case. Another option would be to expand the list of designations to include chemicals known the state of California to cause cancer under Proposition 65, because that is a more comprehensive list which includes chemicals identified as carcinogens by EPA, IARC, NTP, FDA, and NIOSH, as well as the state's qualified experts.

I also note that the use of these designations is not consistent in the document. On page vii, chemicals in the EPA A, B, C or "known/likely" classifications are considered carcinogens, whereas on page 10 only the A, B and "known/likely" categories are named. Given the fact that the "C" classification refers to a "possible" human carcinogen I assume that the authors intended to include only the "A, B and known/likely" classifications. This raises an interesting point, however, with regard to using expedited approaches to encourage testing for the purpose of filling data gaps. It would be beneficial to include as carcinogens chemicals that fall into the categories that designate "possible" carcinogens or chemicals with "limited evidence" of carcinogenicity (such as the "C" classification from US EPA and the "3" classification from IARC) as a means of encouraging the completion of sufficient studies to properly classify these chemicals as to their carcinogenicity. Chemicals falling into these categories would need to be screened first, to confirm that the reason for the designation is that insufficient/limited data are available on carcinogenicity. Certain chemicals with these classifications may actually have a sufficient data set that provides only limited evidence of carcinogenic effects. In these cases, rather than treating the chemicals as carcinogens, an additional safety factor could be applied to an interim noncancer benchmark to address the limited carcinogenic concern. In other cases where the classification is based only on a limited data set the chemical could be classified as a carcinogen and an interim cancer benchmark derived using the limited data. Alternatively, all chemicals in these classes (e.g., EPA "C", IARC "3") could be treated in the same way, that is to derive an interim noncancer benchmark and apply an additional safety factor to protect for carcinogenic concerns.

Fourth, the logic of the derivation of the interim cancer benchmark is problematic based on how it is generally described in the document. The approach hinges on whether or not a cancer benchmark study is available, with other options to follow if no adequate benchmark study can be identified. One of the secondary options is to use a TD50 to derive a CSF based on a regression. If a TD50 is available, then a cancer benchmark study must necessarily also be available, meaning there is no need to use the TD50 to derive a cancer potency. True, the

study may not meet the criteria laid out in this document, but application of the linearized multistage model to derive a CSF is still possible. I suggest that applying the LMS model to a limited study would be more acceptable and closer to results expected under the traditional method than use of the TD50/CSF regression would be. The decision taken by Hoover et al. to use the LMS model rather than using the relationship between TD50 and CSF to derive potencies was in part a result of feedback from the scientific peer review of the method.

> 4. Do you believe that, given the goal of the expedited methodology, these approaches [involving limited use of chemical analogs] are scientifically defensible or should be expanded to include additional techniques and constituent groups?

The use of strong acid salts in place of the parent compound has been widely applied based on the presumption that in either case the same active form is present. Similarly, use of TEFs for dioxins and PAHs is also widely accepted. I agree with the approach to use expanded QSAR on a case by case basis.

> 5. ...provide recommendations for an expedited review process that you believe would meet reasonable standards for scientific rigor.

I agree that expediting the review process is critical and any number of options in this regard would be acceptable; I provide a few general recommendations as follows. To simplify the process of review, I recommend that documentation on each chemical be kept minimal with the main focus of the review being the methods (the feasibility of minimizing the documentation will depend in part on the final selected approach). I suggest that the method and results first be internally reviewed within the relevant EPA program office related to the HWIR. A quality control process should be employed to ensure the accuracy of the calculations. The method and results could then be provided to a small panel of outside experts for review and comment. Review and comment should be provided in writing. If required, the experts could meet with a facilitator present to resolve contentious issues. Public comments, again with a focus on the method, could be solicited through a notice and/or a public meeting. In terms of detailed comments from the public on a specific chemical, the main avenue of addressing such comments should be through an in depth review of that chemical by the agency upon request by an interested party. The interested party would have to provide a written justification for the request for an in depth review, including relevant studies not already considered by the agency. The program office could choose to proceed with the review if the request is supported by a reasonable justification (guidelines defining "reasonable justification" would need to be determined). The program office would have some time period over which such a review could take place or could also defer to another program office that has already identified the chemical as being under review or has scheduled the chemical for review. In either case, the interim benchmark would be retained until the in depth review is completed. In this way, the goals of the expedited regulatory process would be met and at the same time stakeholders' concerns would be addressed.

CANCER ISSUES FOR REVIEW

1. For constituents for which one or more dose-response carcinogenicity studies exist, the LMS model will be used to estimate the CSF...Do you agree with the proposed use of the R factor to account for the uncertainty inherent in an expedited approach and, if not, please recommend alternatives for consideration by OSW.

I do not agree with the use of an adjustment factor. First, the expedited approach applied in Hoover et al. is NOT the same as the approach being applied in the current document. The approach in the current document is virtually identical to the traditional approach and if this approach is retained it would be inappropriate to apply a factor to adjust the results (see page 13 of these comments for more discussion on this point). Second, even if the Hoover et al. approach were adopted, the adjustment factor would still be unnecessary. On average, the expedited potencies are more conservative than the conventional potencies not less. There are a few cases where the expedited potency is significantly lower than the conventional; these outliers were explained in Hoover et al. There are also cases where the expedited potency is significantly higher than the conventional one. Applying an uncertainty factor would be unnecessarily conservative. Finally, Hoover et al. found that the differences between the expedited versus conventional potencies were comparable to the differences between cancer potencies derived by California versus EPA. Thus it applying the "R" factor to adjust expedited potencies would be comparable to applying a factor to traditionally derived potencies as a means of addressing uncertainties related to differing scientific judgment between agencies. Such an adjustment has not been considered necessary.

Questions 2. and 3. (answered together)

2. Do you agree that [using a regression analysis to relate TD50s to CSFs] is an appropriate technique to derive interim cancer toxicity values?

3. The use of the TD50/CSF relationship is a quicker less resource-intensive method. Do you think that these advantages are sufficient to justify a change in the methodology such that this method is used in lieu of the linearized multistage model for constituents that have quantitative cancer data?

As noted above, if a TD50 is available, then a cancer benchmark study is available and there is no need to use the regression relationship. I do not agree that the TD50/CSF is significantly less resource intensive than applying the LMS model. It is a simple matter to identify appropriate studies from the CPDB and run the default model and, as discussed in the Annex to Hoover et al., this approach produces more reliable potency estimates that are consistent with the traditional approach compared to potency factors derived based on TD50s.

ADDITIONAL COMMENTS ON CANCER ISSUES

If the original approach is retained, in which only chemicals already identified as carcinogens by EPA, IARC or DHHS are retained for interim cancer benchmark development, I think it is unlikely that a cancer bioassay would NOT be available because such evidence would have been required as part of carcinogen identification. Thus, the contingencies provided for chemicals without adequate cancer bioassays seem unnecessary (the exception of chemicals identified as carcinogens based strictly on structure-activity relationships is discussed in the next paragraph).

Beyond the issue of locating adequate cancer bioassays, I would also predict that for most carcinogens identified by EPA/IARC/DHHS cancer potency factors would in fact be available already. This would be particularly true if Hoover et al. is used as a secondary source of potencies as so indicated in this document. Chemicals identified as carcinogens based on strong structure-activity information (e.g., dioxins, PAHs) are likely to have TEFs and therefore cancer benchmarks already.

I find that the currently proposed method is not sufficiently distinguished from the traditional approach to offer gain in terms of resources and time. I support instead an approach which involves an expedited determination of carcinogenicity where needed and application of the expedited method of Hoover et al. The Hoover et al. approach could be modified slightly by including an updated literature search to check for more current studies that are not available in the CPDB. The addition of the literature search would compromise the expedited aspect to some degree but the benefits of having a more complete database may offset the disadvantage in added time/resources. The Hoover et al. approach has already been peer-reviewed, both by expert panels in the context of regulatory review and by scientific peers in the context of publication in the literature. An approach was laid out for study selection, which is a critical aspect of expedited potency derivation. Given the goal of carrying out the HWIR process within a time frame of roughly two years (based on information in the charge to reviewers), I recommend against attempting to develop a new "expedited" approach and instead capitalizing on the efforts of the state of California.

NONCANCER ISSUES FOR REVIEW

1. Do you believe that the use of probability distributions is an appropriate alternative to the traditional uncertainty factors given the limitation of the supporting data set?

Generally speaking, a probabilistic approach is designed to be more realistic and less conservative than a default deterministic approach. The use of probabilistic uncertainty factors would not necessarily address the quantification of uncertainty related specifically to the expedited aspect of the approach. The document also points out that using the probabilistic uncertainty factors could have the effect of making the expedited RfDs/RfCs less conservative than the traditionally derived values.

As discussed in the "Highlighted Issues" section, I am in support of the use of probability

distributions to represent uncertainty factors along with other improvements to the derivation of toxicity reference values. I also note that under HWIR, probabilistic techniques are applied in the exposure assessment and use of probabilistic techniques on the dose-response side would therefore be appropriate. Uncertainty in dose-response can be more significant than uncertainty in exposure (see Hill and Hoover 1997 for a case study illustrating this point). However, given the goals of an expedited philosophy and the evolving nature of the field, it may not be practical to institute this aspect of the approach at this time. It may be more appropriate to apply the more sophisticated and realistic methods during an in depth review, in which toxicity reference values would be derived to replace the expedited interim values. I suggest that OSW evaluate the advantages of using these methods for deriving the interim benchmarks against the potential disadvantages of the increased time and resources required, as well as the likely increased difficulty in obtaining consensus on exactly what approach to use in place of the default approach.

2. Regression of TLVs vs. RfCs

I think this is a valid expedited approach. The OSW document should, however, point the potential flaws/uncertainties in this approach. The TLVs have been criticized as being insufficiently health protective (see for example Roach and Rappaport, 1990) and this should be noted. Given the fact that the approach uses a lower 95% confidence bound on the regression, however, I do not consider the potential underconservative nature of the TLVs as a significant issue.

3. Use of representative distributions for RfDs and RfCs to establish interim benchmarks for data poor constituents.

I support the use of current distributions to estimate an RfD/RfC for constituents without data, and have recently applied a similar (though less conservative) approach in a risk assessment of persistent compounds in Canadian breast milk (Hoover, 1999). I think the use of the 5th percentile is sufficiently conservative for an interim benchmark, and is most likely overconservative (as expected based on the definition of the 5th percentile).

4. Grouping chemicals on the basis of class and using those distributions of RfDs/RfCs to select interim benchmarks.

In principle this is a good idea. However, given the well known difficulties in using available data on a chemical class to predict toxicity for a new member of the class (i.e., an untested chemical), it may be most straightforward to use the full data set. This is particularly true given the goal of the approach, which is to produce an interim benchmark quickly and within the constraints of schedules and budgets. I have particular concerns with the small groupings (i.e., $n < 10$). The tail of the distribution would not be well defined with such a small data set. I also disagree with the use of a catch-all category for "miscellaneous pesticides". Has an expert in structure-activity reviewed the approach in detail? Have sensitivity analyses been conducted by, for example, using alternative groupings determined by a second expert? One option would be to default to the full data set and deviate from this default only for chemical groupings with large underlying databases and strong evidence of the structure-activity

relationship. Another option would be to derive an RfC/RfD based both on the overall data set and the appropriate chemical grouping and select the more conservative value.

ADDITIONAL COMMENTS ON NONCANCER ISSUES

It is well known that the results of traditional toxicity assessment on can vary widely depending on the assessor's scientific judgment and the data he/she had at the time of the assessment. Vermeire et al. (1999) cited a study by Dourson and Lu of noncancer benchmarks (i.e., RfDs published by EPA and ADIs published by WHO) which illustrated this point (38 chemicals had RfDs within a factor of 3 of the ADIs, 20 within a 3 to 30-fold range, 6 within a 30 to 300-fold range, and 1 that had an RfD that was 700-fold different from the ADI). This study emphasizes the importance of making the criteria for selecting, evaluating, and applying uncertainty factors to the relevant studies explicit and transparent, such that multiple reviewers would be less likely to produce widely divergent results. This would be particularly important in the context of an expedited approach because such an approach does not allow for the luxury of full consideration of all related data.

As discussed in Vermeire et al., one method for clarifying the noncancer risk assessment process is to develop default assessment factors that are backed up by extensive scientific discussion. The context for the current assessment would not allow such a resource-intensive approach, but Vermeire's comments do provide a general framework that could be followed to expedite the review of noncancer studies. It will be crucial to distill the study review process down to a checklist or some such simplified method, in which the reviewer only has to determine particular characteristics of the study in order to proceed through decision points. For example, a reviewer might answer questions related to study duration, availability of NOAEL vs. LOAEL, etc., and use those answers to assign appropriate default uncertainty factors. It may be necessary to assign the initial review of each study to one reviewer, with a series of checks by other individuals, to ensure consistency of the reviews.

I suggest that chemicals to be evaluated for noncarcinogenicity be grouped on the basis of the available databases for the chemicals. An expedited approach will be simple for chemicals with one or two studies, and these chemicals should be reviewed first. For chemicals with extensive databases but no noncancer benchmark, there is likely a complicating issue that has prevented derivation of such a value and this would need to be investigated. Applying an expedited approach to such a database would be more difficult, though certainly still possible.

If a chemical has limited data available, I would advise using that data as the basis for a benchmark and comparing the results to those obtained using a statistical approach. The more conservative value of the two could be selected as the interim benchmark.

When a benchmark is derived using a statistical method (in the absence of sufficient chemical-specific data), physicochemical properties/structure activity relationships should be evaluated to determine whether the chemical may be inferred to be "unusual" (e.g., potential for being unusually toxic). If a chemical is deemed to be potentially unusual, calling into question the applicability of the statistical approach, the interim benchmark can still be adopted but recommendations to carry out additional toxicity studies and an in depth review of the

chemical should also be made.

When a conservative, statistical approach is used to derive a benchmark, it would be useful to conduct a "real world" check of the value to determine the applicability of the benchmark (e.g., use the interim benchmark to calculate risks associated with natural background and/or typical exposure levels, if sufficient data are available). If the real world check indicates that the benchmark derived is potentially overconservative, the interim benchmark can still be adopted but recommendations for additional toxicity studies and an in depth review of the chemical should also be made.

SPECIFIC COMMENTS BY PAGE

I include specific editorial comments below, as well as general comments on approach where relevant. Depending on which approach(es) is ultimately selected, some of the specific editorial comments may not be relevant to the final document.

p. vii. First bullet under carcinogen class make consistent with later reference and change to reflect final adopted approach.

p. vii After bullets, change sentence to "If none of these criteria is met..."

p. viii 4th bullet change to "..the linearized..." (not "linear")

p. ix Clarify cases where the use of the TD50/CSF regression would apply. As discussed above, based on my reading of the document this would only occur when the cancer benchmark study which underlies the TD50 is considered by the OSW criteria to be "inadequate" for dose-response analysis. Or, the intent may be to replace the use of the LMS model with the TD50/CSF regression. I disagree with either approach, but in any case the document should be clarified and made consistent in this regard.

p. x 3rd bullet, change to be consistent with page viii 2nd bullet, i.e., either use "oral or inhalation" OR "drinking water, food, gavage, or inhalation"

p. x. NOAEL is the "no **observed** adverse effect level"

p. x Clarify what other uncertainties are considered to be uniquely related to the expedited approach. Database deficiencies are the most obvious concern, but this uncertainty is also a potential problem in the traditional approach.

p. xi Rearrange sentence after bullets to clarify meaning - here is a suggested revision, but check for appropriate meaning: "Examples include the use of pyrene as a chemical analog for other noncarcinogenic PAHs and the use of one isomer as a chemical analog for a second isomer (e.g., o-nitrotoluene for p-nitrotoluene)."

p. xii First sentence under "Expedited Review Process": Remove "as well" at the end of the sentence (redundant to "In addition" at the beginning of the sentence).

p. xii Third sentence under "Expedited Review Process" - if there are "major points of contention" this would made application of an expedited process even more difficult than a traditional one potentially.

p. 1 Provide more definitions ("waste concentration limit" "exit criteria"), citations, background on HWIR, etc.

p. 1 Sentence before first set of bullets: "... predicted risk**s**..."

p. 1 3rd bullet, delete parenthetical (not needed/informative here).

p. 1 Clarify tiered use of data sources. According to what I found in the HWIR proposed rule, OSW consults IRIS, HEAST and then NCEA (provisional EPA benchmarks). Is the use of ATSDR and Cal/EPA proposed here or has this been endorsed/applied previously by EPA? The way this is written it sounds as though the use of these alternative sources is standard practice. If this were the case or even if this is proposed in the current document, this raises an additional issue related to application of Cal/EPA expedited potency factors. Specifically, the level of data/effort required for deriving a Cal/EPA expedited potency is less than what is proposed for the "expedited" approach in this document. That would mean that a less rigorous value (i.e., the Cal/EPA expedited potency) would take precedence over development of an interim benchmark based on the proposed methodology. As noted throughout these comments, I would support the use of currently available expedited potencies with application of the Hoover et al. approach to fill in missing values. If this proposed revision to the approach is adopted, then the logic problem mentioned in this paragraph would not apply.

p. 1 What is the implementation scheme referred to in the last sentence? Describe and provide appropriate citations.

p. 2 Rewrite last sentence. Current sentence is a run-on sentence and does not appear to say what I believe was the intended meaning. Suggested rewrite: "By taking the approach proposed here, the Agency can calculate risks based on interim benchmarks while leaving open the possibility of changing/refining the interim benchmarks as more information becomes available. The interim benchmarks are designed to be conservative, such that once a refined/final benchmark is derived predicted risks would be lower than those based on the interim benchmark." Shorten and refine as needed.

p. 3 An NTP bioassay report is an original study, not a "secondary data source".

p. 4-5 Use of terms such as "critically evaluated", "adequate" - careful definition of what these terms mean is needed.

p. 6 First full sentence suggested rewrite: "Long-term carcinogenesis bioassays in rodents usually include at least 50 animals per sex in each of three treatment groups and in a concurrent control group. The duration of the long-term bioassays is usually

18 to 24 months."

p. 6 Third sentence, second full paragraph - provide citation for statement that less than 15 days is an acceptable duration for developmental studies.

p.6 Sixth sentence, second full paragraph, add word: "Mortality or endpoints relevant to cancer are not selected...as these **endpoints**..."

p.7 Provide the number of constituents rather than saying "Given the number of constituents..." Are these constituents known? If so, include list of them as an Appendix. This would allow reviewers to suggest data sources and approaches for evaluating these chemicals more effectively.

p.7 What are the regulatory goals? Spell them out or at least provide relevant citation. Providing sufficient background is essential for complete and effective review as well as clarity in the document.

p.7 Last sentence, first paragraph add phrase: "...that formally incorporates consideration of the larger uncertainties that will be introduced **by applying an expedited approach rather than a traditional approach**." As well, define better what the "larger uncertainties" are.

p. 8 Second full paragraph first sentence, suggested rewrite - "...the expedited approaches **proposed in this document**..."

p.9 As noted above, Figure 3-1 is not consistent with the figure on page vi. Revise one or the other for consistency.

p. 10 Ensure the listed classifications are consistent with earlier information in the document (i.e., EPA Group A, B or C, OR EPA Group A or B).

p. 10 Which chemical constituents of concern for the HWIR rule have been identified as carcinogens by EPA, IARC or DHHS but do not yet have a cancer benchmark? I would guess that since the identification process is slow, most of the chemical carcinogens would have a benchmark available from one of the listed data sources.

p.10 First sentence under 3.1.1: What exactly is "a shortened version of the traditional process"? How will the traditional approach be shortened? The literature search does not seem to be any different from the traditional approach - how will it be determined which studies will be reviewed and which ignored in order to achieve an "expedited" process?

p. 11 Second sentence, second full paragraph: Though this sentence states that "one or two" studies for each constituent will be collected and analyzed, no details on exactly how this limited database will be selected are provided. The best way to clarify this approach would be to provide the full list of constituents identified as

carcinogens and work through a few examples in an Appendix. It may be determined for example that there is only one bioassay for many of the carcinogenic constituents that do not yet have benchmarks from EPA or secondary sources. Then the issue of restricting the database becomes moot.

p. 12 Fourth full paragraph, sentence before bullets: This sentence clarifies the fact that RTI did the analysis on the "R" values based on data provided by Hoover et al. This should be clarified in an earlier description of "R" (i.e., that the underlying data was obtained from Hoover et al., but the analysis was carried out by RTI).

p. 13 First full sentence: The process proposed in this document is NOT essentially the same as Hoover et al. Probably the most significant means employed by Hoover et al. for expediting the CSF estimation was the elimination of the literature review. As noted on page 269 of the article "Use of information in the [Carcinogenic Potency Database] can facilitate several tasks that consume considerable time in conventional assessments: performing the literature search, collecting all relevant data sets, excluding poor data from the analysis, and identifying the most sensitive species." In the current approach, the requirement for literature review is maintained making it significantly different from Hoover et al. This is a key difference in terms of interpreting the R value. Some of the explanation for the differences between the CSFs derived by Hoover et al. under the expedited approach as compared to the traditional approach lay in the literature review. For example, in one case the regulatory agencies rejected the most sensitive study available based on knowledge gained in the literature review while the expedited potency was based on this study. In other examples, the studies (including epidemiological) used by the regulatory agencies were not available in the CPDB. If a literature review had been performed, these discrepancies would likely have been resolved and the ratios between certain expedited and conventional potencies would have been closer to 1. Thus, as pointed out above, the use of an adjustment factor based on "R" to CSFs derived using the somewhat restricted traditional approach proposed in this document is likely to be unnecessarily conservative.

p. 14 Existing TEFs have already been applied to derive corresponding cancer potencies in numerous circumstances. Similarly the use of strong acid salts and isomers is accepted. I would consider these as standard, accepted methodologies and don't think it's necessary to highlight them as being part of the expedited approach.

p. 16 First sentence, second paragraph: "Factors considered in selecting the TD50...include...elimination of a potentially time-consuming process for determining if the constituent was a likely carcinogen." This statement is not consistent with other parts of the document, as the only time a TD50 would be used to derive a CSF under the proposed approach would be if the constituent had already by identified as a carcinogen under the "time-consuming" review processes of EPA, IARC or DHHS.

p. 16 Third and fourth sentences - Justifying the use of inverse CSF or URF by stating that "a low TD50 indicates a high potency whereas a high CSF or URF

indicates high potency" is not necessary for a scientific audience. The fact that the TD50 is a dose, whereas the CSF/URF are in units of risk per unit dose makes apparent the need for the inverse variable.

p. 17 Grammar of last sentence first paragraph: ..."however, advantage**s of the regression approach include that**..."

p. 17 General comment: As noted in the Annex to Hoover et al., a simple formula (properly scaled to adjust for differences in the treatment of experiment duration between EPA and Gold et al.) can be used to derive the CSF from the TD50 rather than applying the regression. The formula is best applied when background incidence is negligible but still produces reasonable estimates even when background incidence is not negligible. As discussed in Hoover et al., however, the most reliable approach to deriving a cancer benchmark is to run the multistage model using the cancer benchmark study underlying the TD50.

p. 21 The difference between the approach proposed to derive interim noncancer benchmarks and the traditional approach is not clear. There are no clear guidelines provided to explain how the database of studies to be reviewed for deriving the interim benchmark would be limited.

p. 24 Accounting for incomplete database - how is it proposed to translate this information into an adjustment factor? This is not provided in Table 3-1 (as it was apparently not addressed in Baird et al.).

p. 25 It is not clear why Baird et al. would advance the stated goal of ensuring that the interim benchmark derived using expedited methods would be conservative relative to the traditional RfD/RfC. Similar comment for Swartout et al.

p. 27 I suggest expanding the citations for petroleum hydrocarbon surrogates [see end of these comments for additional references].

p. 27 Why is this discussion of chemical analogs expanded and different from the one provided under the cancer benchmark section? Why is the acetate salt now included, whereas before only strong acid salts were mentioned? I suggest that the entire discussion of chemical analogs and expanded QSAR be taken out of the sections on cancer and noncancer benchmarks and incorporated into one section on this topic.

p. 30 Third sentence - Since the document is dated June 1999, I am not clear on why the data are current only to August 1998. I suggest updating to current database prior to finalizing the document.

p. 30 Second bullet, delete first parenthetical not needed.

p. 30 Second bullet, second parenthetical - Based on previous discussions in the document, it is not the "true RfD" but rather the "RfD derived using traditional

methods." So suggested rewrite:, "...to be 95 percent confident that the new chemical would have an RfD derived using traditional methods at or above the RfD assigned by this expedited method."

p. 30 What is the basis for selecting the 5th percentile? Will this be evaluated/adjusted using input from risk managers?

p. 32 Section 3.2.2.5 - An expert on SAR should be consulted. As discussed previously, sensitivity analyses should be run to consider the impact of alternative groupings. I also suggest carrying through a few case studies as examples.

p. 42. Hoover et al. is incorrectly cited. Title should read "Improving the regulation of carcinogens by expediting cancer potency estimation."

Appendices in general: Run spell check (e.g., p. D-1, D-14, D-15 "halogenated"; E-21, "miscellaneous")

Appendix C: For better ability to review, sort on chemical name rather than RfD/RfC.

SUMMARY OF PROPOSED REVISED APPROACHES

In consideration of my foregoing comments, I recommend the following general frameworks for deriving interim cancer and noncancer benchmarks. Explicit details will need to be worked out and will depend in part on exactly which chemicals need to be addressed under the proposed HWIR. Elements of the review process are not discussed further (see page 5 for recommendations).

Interim Cancer Benchmarks

For chemicals identified as carcinogens by EPA, IARC and DHHS and chemicals identified as carcinogens using an expedited determination, apply the approach of Hoover et al. to derive interim cancer benchmarks. The approach involves:

Use of cancer dose-response data that has been evaluated and extracted from the literature and summarized in the CPDB by Gold et al.;

Selection of appropriate study from the CPDB based on explicit criteria (adopt from Hoover et al. or modify as appropriate to current goals); and

Application of the LMS model to the selected study to derive an interim cancer benchmark.

As an optional addition to this approach, use a current literature search strictly to identify newer studies not included in the CPDB. Studies from the literature could be evaluated using the CPDB approach in order to maintain consistency and selected using the study selection criteria. Strong acid salts could be used in place of parent compounds (as well as other salts provided that the active form is the same). TEFs could be applied as appropriate.

Interim Noncancer Benchmarks

A suggested framework for the derivation of interim noncancer benchmarks is as follows:

Conduct comprehensive literature search on chemicals under consideration for benchmark development. Issue public notices to elicit additional data.

Establish explicit screening criteria in the form of checklists for reviewers.

Sort chemicals based on extent of databases. E.g., establish groups of chemicals with no adequate studies, groups with 1-2 adequate studies and so forth.

Note chemicals that were found to have recent, adequate cancer bioassays that have not been reviewed by EPA, IARC or DHHS under a carcinogen identification process. Use current EPA carcinogen identification guidelines to screen those studies and determine whether an expedited determination of carcinogenicity is warranted. If so, consider this chemical for development of interim cancer benchmark following approach described above.

Note chemicals that are classified as EPA Group "C", IARC Group "3" and other comparable categories.

Beginning with chemicals having small databases, evaluate studies using checklist approach. Based on the checklists prepared by reviewers assign appropriate uncertainty factors. Include uncertainty factor to address limited carcinogenic concerns as appropriate.

For chemicals with extensive databases [to be defined], call on expert reviewers familiar with these chemicals to determine the reason for no benchmark being available and the practicality of applying the expedited approach to such an extensive database. Use criteria established for selecting adequate studies and checklist approach to defining uncertainty factors if considered appropriate (e.g., a chemical may have extensive pharmacokinetic data indicating that the default expedited approach would not be appropriate).

For chemicals with limited studies and none considered adequate based on screening criteria, derive interim benchmarks in two ways. First, use the limited studies and the default expedited approach to derive a benchmark. Second, use a statistical approach (e.g., select a value based on the full distribution of RfDs/RfCs) to derive a benchmark. Compare the two and select the more conservative benchmark. This approach will encourage the conduct of additional toxicity studies in order to fill the data gaps and reduce the conservatism of the interim benchmark.

When applying a statistical approach in the absence of any chemical-specific studies, evaluate the properties of the chemical to determine whether unusual toxicity might be expected. If natural background data or typical exposure levels of the chemical are available, use the statistically derived interim benchmark in combination with the exposure data to estimate risks. Based on either of these checks, a recommendation to conduct an in depth review and/or initiate toxicity studies could be made. The interim benchmark derived based on the statistical approach would still be applied until a more detailed evaluation could be

conducted.

REFERENCES

Some of the references below were cited in my comments (I have not listed references already cited in the main document such as Hoover et al., Baird et al., Swartout et al.). I also include additional potentially relevant sources published in 1999 that I identified using a citation search (based on references cited in the document). References for the use of surrogates in assessing petroleum hydrocarbons are also provided.

Brand, K.P., Rhomberg, L. and Evans, J.S. 1999. Estimating noncancer uncertainty factors: Are ratios of NOAELs informative? Risk Anal 19(2):295-308.

Carlson-Lynch, H, Price, P.S., Swartout J.C., et al. 1999. Application of quantitative information on the uncertainty in the RfD to noncarcinogenic risk assessments. HERA 5(3):527-546.

Hattis, D., Banati, P., Goble,R. and Burmaster, D.E. 1999. Human interindividual variability in parameters related to health risks. Risk Anal 19(4):711-726.

Hill, R. and Hoover, S. 1997. Importance of dose-response model form in probabilistic risk assessment: A case study of health effects from methylmercury in fish. HERA 3(3):465-481.

Hoover, S. 1999. Exposure to persistent organochlorines in Canadian breast milk: A probabilistic assessment. Risk Anal 19(4):527-545.

Hutcheson, M.S., Pedersen, D., Anastas N.D. et al. 1996. Beyond TPH: health-based evaluation of petroleum hydrocarbon exposures. Regul Toxicol Pharmacol 24(1):85-101.

Roach, S.A. and Rappaport, S.M. 1990. But they are not thresholds - A critical analysis of the documentation of Threshold Limit Values. Am J Ind Med 17(6):727-753.

Roberts, S.M. 1999. Practical issues in the use of probabilistic risk assessment. HERA 5(4):729-736.

Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG). 1997. Selection of Representative TPH Fractions Based on Fate and Transport Considerations. TPHCWG Working Group Series, Volume 3. Prepared for Association of American Railroads, United States Air Force, and the TPHCWG. Amherst Scientific Publishing. Amherst, MA.

Total Petroleum Hydrocarbon Criteria Working Group (TPHCWG). 1997. Development of Fraction Specific Reference Doses (RfDs) and Reference Concentrations (RfCs) for Total Petroleum Hydrocarbons. TPHCWG Working Group Series, Volume 4. Prepared for Association of American Railroads, United States Air Force, and the TPHCWG. Amherst Scientific Publishing. Amherst, MA.

Vermeire, T., Stevenson, H., Pieters M.N. et al. 1999. Assessment factors for human health risk assessment: A discussion paper. Crit Rev Toxicol 29(5):439-490.

ATTACHMENT B

Comments on*"Conceptual Approach to Establish Interim Human Health Benchmarks"*

Daland Juberg

ICTM

Reviewer Comments and Critique on:

"Conceptual Approach to Establish Interim Human Health Benchmarks"

Peer Review Draft – June 1999

RTI Project Number 7200-007

A. Reviewer Responses to *General Issues for Review*

Q. *Do you feel that the major sources cited (e.g., Hoover et al., 1995) in this report have gained sufficient acceptance in the risk assessment community for application in a regulatory context?*

A. This reviewer believes that the risk assessment community is not sufficiently proficient and knowledgeable in the uses and limitations of some of the major approaches (e.g., R factor) in this Draft to warrant application in a regulatory context (without some additional discussion). Some of the other citations (Baird, Swartout) are more common within this context.

Q. *Are there alternative methods not presented in this report that you would recommend for use in deriving interim health benchmarks?*

A. Presently, there are various methods and approaches for prioritizing chemicals in terms of toxicity (and bioaccumulative capacity and persistence within the environment), but there are fewer widely known, used, and accepted methods for deriving health benchmarks. The current approach described in this Draft could be appropriately used as a first tier, screening method for purposes of putting some type of framework around hundreds of constituents.

However, and very importantly, a chemical-specific subsequent review or check must be conducted to determine whether the approach described in this Draft has appropriately characterized the chemical's toxicological properties. For example, if the Draft approach were used and an interim benchmark determined for a constituent, and yet a followup check determined that the value was lower than current natural background levels, then the methodology and interim benchmark would have to be revisited as the approach would have been ineffective in its intent and purpose. If many or all values developed within this approach turn out to be very low values, those approaching background values, then the approach will not be effective as a screening tool.

It is also recommended that following development of screening values, there be some mechanism for conducting a site-specific review, based on the intent and interests of the petitioning party. In other words, if an industry or owner of a property could demonstrate through risk assessment or other viable techniques, that there are no human exposure or receptor pathways, then the developed interim benchmark may be increased (less conservatism) to a mutually agreeable level.

Do you support the logic illustrated in the Figure 3-1 hierarchy for development of interim benchmarks?

Two major changes or proposed additions. (1) Following development of the interim benchmark, there needs to be some provision and mechanism for allowance of external peer-review and this should be included in the diagram. (2) Also, there needs to be some flexibility and chemical-specific provisions for what constitutes a carcinogen. Our knowledge in cancer biology has advanced sufficiently (although the major cancer classifying mechanisms such as IARC, IRIS, etc. have not always kept sufficiently up to date in terms of classification) that we now know there exist certain species-, mechanism-, or chemical-specific factors that preclude the direct extrapolation of animal cancer results to humans. Just because a chemical has been demonstrated to be positive in a rodent cancer bioassay and is labeled as a probable or possible human carcinogen is not sufficient cause for moving it through the hierarchy as proposed and setting an interim cancer benchmark. Half of all chemicals tested in standard high-dose animal cancer bioassays, whether occurring naturally or produced synthetically are "carcinogens"; there are high-dose effects in rodent cancer tests that are not relevant to low-dose human exposures and which contribute to the high proportion of chemicals that test positive (Ames and Gold, 1997). This reviewer believes that development of interim cancer benchmarks needs to be evaluated very carefully so we do not erroneously develop such for a chemical that truly has no human carcinogenic potential.

Do you believe that, given the goal of the expedited methodology, these approaches are scientifically defensible or should be expanded to include additional techniques and constituent groups?

They should not be expanded at this point until additional information is available to warrant their use. This approach needs to be very careful with the use of equivalency factors, and perhaps even more careful with the use of sulfate or salt forms of a chemical as a surrogate. This should only be done on a chemical-by-chemical basis.

Please provide recommendations for an expedited review that would meet reasonable standards for scientific rigor.

Process review must include external (to the Agency) scientists. In addition, the review must include some mechanism for insuring that a developed benchmark is achievable within the environment, that it is not below background or natural levels (e.g., metals, naturally occurring PAHs), and that it is not overly conservative so as to decrease the utility of this approach as one that is effective and important in development of human health benchmarks. Finally, the review process should require that once a benchmark is developed, that a review of the database for the specific chemical be scanned or reviewed to insure that the database is consistent and supportive of whatever benchmark is proposed.

B. Reviewer Response to *Cancer Issues for Review*

Do you agree with the proposed use of the R factor to account for the uncertainty inherent in an expedited approach and if not, please recommend alternatives for consideration by the OSW.

The R factor is not a commonly known or used entity in the risk assessment arena, particularly

in the establishment of environmental health benchmarks. One comment on the use of the LMS model – while this has traditionally and historically been the model of choice for the Agency, there has been a recent shift in scientific consensus among some scientists and cancer biologists to suggest that a threshold exists for some cancers and that the default use of this model (LMS) may be inappropriate in some situations. Much research has been conducted on certain chemicals, such as formaldehyde, chloroform, to name a few, that has advanced greatly our understanding of the carcinogenic potential and process, and for chemicals such as formaldehyde, cancer risks have been dramatically and greatly reduced from previous estimates using the LMS. The salient point is that we must continue to use all of the available scientific knowledge during the establishment of any human health benchmark. If the R factor is going to be used as a default measure to account for uncertainty in this approach, there needs to be a validation of the value derived (using the R factor) against background or natural levels in the environment to insure that this factor is not overly conservative and has resulted in an interim health benchmark that is below ambient levels.

Do you agree that regression analysis is an appropriate technique to derive interim cancer toxicity values?

Obviously regression analysis and evaluation of the TD50s and CSFs through this technique are purely statistical methods. There needs to be some biological meaning and toxicological analysis incorporated in any approach. This technique may be appropriate as a first-tier screening method, but again, should be followed with more biologically meaningful in-depth analysis depending on the data and time available.

Do you think that the advantages of using the TD50/CSF relationship (vs. the LMS) are sufficient to justify a change in the methodology such that this method is used in lieu of the LMS for constituents that have quantitative cancer data available?

Perhaps. What needs to be done is a comparative analysis for several chemicals, evaluating them through both the LMS and TD50/CSF approach. It is important to determine what the interim values are through both approaches and if one model yields consistently higher or lower values than the other. Obviously with only two models, it would be difficult to determine which one is closer to a true health benchmark, but this could be evaluated by retrospective analysis of the toxicological database for each chemical. However, this reviewer would support an alternative to the LMS, where appropriate, particularly if it

removes some of the unwarranted conservatism inherent in a non-threshold model such as the LMS.

C. Reviewer Response to *Noncancer Issues for Review*

Q. Do you believe that the use of probability distributions is an appropriate alternative to the traditional uncertainty factors given the limitations of the supporting data set?

A. Again, there needs to be some comparative analysis done on several different constituents to determine how various interim benchmarks are developed in terms of their end value. The use of probability distributions may serve as a first-tier approach, which could then be followed with the use of traditional UFs which are more constituent specific (e.g., use of various UFs depending on extrapolation from a subchronic to chronic, interspecies sensitivity, etc.). Historically, the traditional approach and use of UFs has appeared to work rather well, although conservatively, and thus unless probability distributions offer distinct advantages, the traditional approach should still be used. It also has enjoyed refinements in its application in recent years and this progress should be included in any application of UFs in this approach.

Do you consider this (regression of TLVs against RfCs) an appropriate use of these data and can you suggest improvements for this approach?

This is purely statistical treatment of evaluating and attempting to correlate values for 8-hr occupational exposure with lifetime permissible inhalation benchmarks. I would be very careful and selective in using this regression technique for this purpose.

From a technical standpoint, do you agree that this approach achieves the desired goal of the methodology, namely that the interim benchmark is equal to, or more conservative than, the traditional benchmark at a specific level of confidence?

Yes, technically, the approach probably achieves the goal, but I believe the goal needs to be

reexamined. The primary goal should be to develop a robust, time-efficient, and scientifically accurate and defensible methodology for evaluating and developing human health benchmarks for exit concentration criteria, among other applications. The goal should not be to maintain conservatism in a particular approach, but do use science and scientific principles to our advantage in applying such knowledge to issues of public health.

Do you think that the use of SAR in this scheme is an improvement to the use of distributions given the uncertainty associated with chemical/physical properties and toxicity mechanism of action? Do you agree with the chemical classes developed to refine the distributions?

For individual chemical classes, the use of SAR is relevant and appropriate. SAR should not be used for all of the constituents collectively, and chemical class independence should be maintained. Finally, I would not place an overemphasis on the use of SAR within this overall approach. It remains a tool for analysis purposes, but should not be used as a stand-alone default method for evaluating constituents and developing health benchmarks.

D. Reviewer Summary Critique and "Large Picture" Comments

If it is not obvious in the following page-by-page comments, it will become rapidly apparent that this reviewer believes there is an excessive emphasis on statistical treatment of data related to development of this interim approach and conversely, not enough biological plausibility and toxicological consideration given to the methodology. The authors will need to defend this position and explain to the users of such an approach, why use of statistics, R factors, distributions, and log functions are so critical at the exclusion of other more biologically meaningful parameters.

This approach and proposed methodology may be used for a variety of purposes associated with OSW activities and responsibilities and it would be very helpful to have some background in the Draft document pertaining to such, as this will help users and reviewers

in knowing what types of exposure scenarios humans might encounter. This type of information is critical and relevant in the subsequent development of appropriate health benchmarks. For example, in setting an occupational standard, the route of exposure is frequently inhalation and the temporal consideration is usually 8-10 hrs/day. For purposes of better developing the proposed approach in the Draft, it is imperative to identify the range and types of exposures that humans may encounter – Will exposures most likely be through inhalation, oral ingestion, or dermal contact? Will exposures be transient in nature or lifetime chronic exposures? If these interim benchmarks are used for exit criteria associated with an industrial facility, it is unlikely that humans will ever have chronic exposures to these various constituents. The approach that is developed and used needs to mirror, to the extent possible, those most likely anticipated human exposures and scenarios. In this regard, the methodology needs to anticipate and forecast the exposures (inhalation, oral, dermal) and durations (minutes, hours, lifetime). Such will be critical in a final analysis as to how conservative each interim benchmark needs to be. Additionally, the OSW should provide some background information on its various regulatory programs and activities, such that users of this methodology will know the basis and impetus for development of such an approach.

A rather large portion of the Draft pertains to the Appendices illustrating the statistical treatment (e.g., log RfDs, etc.) of all the constituents under potential review and the appendices appear to be disconnected from the text. In other words, there needs to be additional explanation or linkage in the text as to what and why the appendices are included. There should be explanation within the appendices themselves as to how the statistical treatment of the constituents is related to the proposed benchmark establishment process. Give examples. Finally, similar to a previously mentioned comment, the primary focus should be on the text and on explanation of the process; minimal treatment should be given to exhaustive presentation of the constituents (in the appendices) at this stage. The detail given to the appendices would be have been better served by showing examples of how the process may or may not work.

There needs to be some mechanism within the proposed approach for provision of external peer review and the ability of non-Agency scientists to participate, either directly or indirectly, in the development and adoption of interim benchmark values. Historically, the Agency has limited participation from external parties, yet there is positive benefit to having multiple stakeholders and interested parties participate in this process.

There needs to be some provision for "field-testing" or validating this approach prior to any further development or promulgation. While the approach has some merit as a screening

tool, there needs to be evaluation and validation of this approach which could be accomplished by taking several different constituents and developing health benchmarks based on the contents of this document. If the resulting values appear far below either natural, background, or detectable ambient levels, then some adjustments within the methodology are required. Clearly, however, there needs to be some evaluation and validation of "test case constituents" before the methodology is further developed.

It is somewhat troublesome that a primary goal of the approach is to insure that any developed interim benchmark is more conservative than traditional health benchmarks. This puts the onus of benchmark development on preserving a policy goal and not allowing the science and knowledge drive the process. This will become quite apparent to users (and abusers) of this approach and may lead to a less than desired acceptance of the methodology by the regulated community. This reviewer would advocate using sound scientific principles and data and letting the "chips fall where they may." As we have learned over the years, most regulatory health benchmarks are very conservative by their nature and intent, and thus, here is an opportunity to shift the policy to one that is not so conservative and more in line with the scientific evidence.

This approach and methodology will be of limited use to most risk assessors who have not had extensive statistical or probabilistic training. Much of the discussion is beyond what many scientists can understand without additional training and education, and given the fact that this is an interim benchmark approach, the level of statistical sophistication is excessive.

E. Reviewer Page-by-Page Comments

Page v – Within the first paragraph, the text should state that the proposed approach is purposely conservative.

Page v – 2nd paragraph – Text might include a sentence that removes any question as to why human health benchmarks typically do not consider or establish values for the dermal route of exposure. To a lay reader, they may question why this route is not included.

Page vii – "In the absence of a constituent……..since it has not been sufficiently established that the constituent is a carcinogen." Good point here and this reviewer commends the authors for restraining any development of a cancer benchmark when there is absolutely no evidence that a constituent has carcinogenic properties or propensity.

Page viii – "The cancer benchmark study must include two or more doses, with the results showing a statistically significant dose-response relationship." What type of dose-response relationship? One that considers only malignant tumors? Combined malignant and benign tumors? Please specify.

Page viii – "If several dose-response carcinogenicity studies exist, the study showing the most sensitive dose response will be selected." At all costs and despite perhaps, other more meaningful data and information? This seems to place greater emphasis on insuring the conservative nature of the proposed approach at the expense of perhaps valuable and meaningful knowledge or information that might enhance the accuracy and defensibility of a derived value.

Page viii – "The linear multistage model will be used to estimate the CSF." This reviewer simply does not agree that the LMS should continue to be used without question, simply because it has been the Agency's choice historically. There is increasing evidence (and

consensus development as noted by a hand vote at the 1998 Society of Toxicology Annual Meeting) that a threshold approach may be more appropriate for many "suspected carcinogens" and that the use of the LMS to evaluate and establish CSFs should be implemented on a case-by-case basis. Be careful with the continued use of a default approach that is decreasing in its utility and favor amongst cancer biologists and toxicologists.

Page viii – 1st paragraph – The whole topic and discussion on the use of the R factor appears non-chemical specific. In other words, again, this approach emphasizes the statistical treatment of constituents and places less importance on the biology of a constituent, the true measure by how and what mechanisms it may impart carcinogenicity. There needs to be some recognition or provisions that permit chemical-specific information to be considered in deriving interim health benchmarks. In addition, there needs to be additional discussion and explanation in this paragraph as to how the R factor will be applied.

Page ix – "Expanded QSAR" – Given the already tentative reliability of some of the proposed approaches for establishing interim benchmarks, this reviewer sees little utility or need to incorporate "expanded QSAR" into this approach. This exercise seems too far removed from other factors (e.g., mechanism, biological meaning, genotoxicity) that have more utility in predicting a constituent's ultimate toxicological properties and effects. Minimize the discussion on this topic. Should only be used in extreme cases.

Page x – 1st paragraph – Do the authors mean an "interim RfD or RfC" will be developed? This is an expedited approach and this reviewer assumes that an interim benchmark will be developed and not a traditional RfD or RfC which would require much more time and peer review. Please clarify.

Page x – Should the benchmark study only need to include two doses? Why not require three?

Page x – "The results should demonstrate statistical significance….." The results should actually demonstrate effects that are toxicologically and biologically meaningful. Of course, statistical significance is important, but any effect resulting from treatment must be important from a toxicological perspective. Additionally, it is very important that there be some degree of concordance between the effect and results of clinical chemistry and pathology evaluation.

These additional observations and data are often confirmatory that an observed change or effect has toxicological significance.

Page x – "For constituents……for the most sensitive or most relevant species." Place greater emphasis on the most relevant species. There is enough conservatism built into this process and approach that the most sensitive species does not necessarily need to be considered as the best choice. Additionally, I would place significant emphasis on selecting a study and/or animal model that is the most relevant TO HUMANS. This may be implicit in this discussion, but should be emphasized.

Page x – 2nd full paragraph – "However, use of an expedited method adds to the uncertainty." Not necessarily in all cases. This sentence is too declarative and cannot stand alone without additional support. Please revise.

Page xi – "Expanded QSAR" Again, this is a stretch and not very useful for this discussion or approach. Minimize or delete.

Page xi – Several comments on the preferred statistical approaches:

"Derivation of an interim RfC from a TLV." Remember that this involves extrapolation of an 8-hr benchmark to a lifetime chronic exposure (e.g., RfC).

"Derivation of …….from the complete distribution….." Again, this is non-chemical specific. Need to confine to chemical class.

Page xii – "In other words, chemicals with interim……." These last three sentences in this paragraph are somewhat confusing. Can the authors please clarify the meaning of this discussion.

Page 1 – "Human health risk analyses performed by OSW require that at least one health benchmark be available….." What if the one that is available is an oral benchmark and inhalation is the only route of concern? Or vice versa? There needs to be some discussion on how to treat this scenario, since it likely will be quite common.

Page 1 – 2nd full paragraph – There should be provisions for also utilizing non-Agency peer-reviewed literature and/or data, particularly if scientific consensus exists for a particular mechanism, effect, etc. The field of risk assessment and safety evaluation has been hampered somewhat by the lack of consideration and use of non-Agency conducted studies and research. Often, these information sources are in fact, of higher quality, since they incorporate GLP, must be conducted according to current protocols and standards, as well as withstand peer review. In contrast, many Agency-produced documents and data sources are not subjected to external peer review and as such, their utility is limited for an approach such as this. This point cannot be stressed enough that a weight-of-scientific evidence approach and incorporation of all relevant data be implemented.

Page 1 – 3rd paragraph – What is meant by "secondary literature"?

Page 1, last bullet – "Develop an implementation scheme that is not directly based on benchmarks." What is the scheme then based on?

Page 2, 1st bullet – Good and appropriate point – "…include additional constituents as EPA-approved benchmarks become available."

Page 2 – "….will require greater use of extrapolation between constituents and species and routes of exposure…." There should be minimal, if any, extrapolation between routes of exposure, given the known differences in metabolism, portal of entry effects, mechanisms, that occur via different exposure routes. There should be a limit as to when, how, and by what means route-to-route extrapolation is conducted.

Page 2 – "A guiding principle adopted…..should be conservative…." NO. This is not a principle, but rather a policy and this should be stated. A principle is a basic truth, law, or

assumption. A policy is a prescribed or selected option. Please clarify as science policy needs to be distinguished from fact based on scientific evidence.

Page 2 – "these interim health benchmarks may change as more information…." This is not likely to happen as we have witnessed through the IRIS program. It is extremely difficult, to the chagrin of the scientific community, to change or alter health benchmarks once they are placed in the public domain. I would not lead the readers of this document to the hope that interim benchmarks are likely to be readily or easily changed, unless the authors can provide some justification and proposed mechanism for such.

Page 3 – Again, what is meant by "secondary sources"? Are these peer-reviewed studies? Agency data? CBI data, case reports? There needs to be some standard by which a study or data are evaluated and used. This approach cannot use just any data that exists.

Page 4 – 1st full paragraph – "A literature search….." Good to identify these various resources and again, you should note that non-Agency research and data should be considered as valid sources of information if they have been peer-reviewed and are considered high quality data and studies.

Page 4 – Good notation of the criteria required for studies that are to be considered for use in risk assessment. However, a requirement should also be added that specifies that a relevant route of exposure (to the human situation) be used if available.

Page 4 – "Exhibit a dose-response relationship" This should not be an explicit requirement, since a well-conducted and high quality study may not, in fact may infrequently, demonstrate a dose-response; yet, these data may still be very useful and important from a risk perspective. Be careful about requiring a dose-response demonstration at the expense of neglecting a better study that is useful, yet did not demonstrate such. There still is merit in such a study for safety evaluation and risk assessment purposes.

Page 5 – Add a bullet that notes the utility and benefit of understanding toxicological mechanism of action.

Page 5 – 1st full paragraph – Excellent point about the mechanisms in various animals that we have learned about and which are not appropriate to humans. These should not be used as a basis for a human health risk assessment. This is the type of current information that needs to emphasized throughout this document. Other examples include consideration of the 1996 revised cancer RAGS and the use of PB/PK information when available.

Page 6 – "benchmark study must be conducted on humans…." Rarely are human studies prospectively conducted and I would revise/rephrase this sentence.

Page 6 – "If several cancer benchmark studies meet the above criteria, the study indicating the highest cancer potency is selected." Again, at all costs and regardless of the utility and perhaps more toxicological relevance of other studies? Does this position hold, regardless of the quality of the study? Again, this statement suggests that greater emphasis is placed on policy and conservatism rather than on scientific merit, rationale, and meaning. I don't think the Agency wants to convey the message that this approach is simply a statistical exercise and is placing greater meaning and weight on those studies that yield the most conservative data with which to derive an interim benchmark. This reviewer cannot support this statement from a risk assessment perspective.

Page 6 – "A noncancer benchmark study should be of chronic……" Actually, an appropriate study length will depend on the type of study conducted and for what endpoint of concern. A minor point, but a benchmark study may not necessarily be of chronic or subchronic duration, but still be used if it evaluated the toxicological endpoint of concern. This is well illustrated in the subsequent discussion about the use of developmental toxicity studies that are not of chronic duration. Please reflect this good example in the beginning sentence as well.

Page 7 – 2nd paragraph – "The traditional….. the assumption that noncancer effects have a threshold whereas cancer effects do not." See comment referring to page viii above. Again, this view is changing as there is a trend developing among many scientists that cancer effects may be threshold-related as well.

Page 7 – 2nd paragraph – "For cancer effects, no exposure level is considered risk free." Again, same comment as above. There needs to be some discussion that brings to light the

changing picture with regard to cancer effects and the fact that a no-threshold assumption may not be correct. More importantly, while the 1996 EPA cancer RAGS are cited, there should be some attempt to weave these into the discussion and into the interim approach proposed.

Page 7 – 1st bullet – This is the first mention of a constituent being absorbed by the dermal layer. Do you want to discuss this route of exposure when it has not been considered previously?

Page 10 – 1st paragraph – It is difficult to discern why in 1999, the linearized multistage model is once again selected as the basis for estimating and developing CSFs and URFs. This seems counterproductive, even for an interim screening approach, to the work that has been done in the area of cancer biology in the past decade and before.

Page 11, 2nd full paragraph – "….some subset of those studies (one or two…..)" Do these "studies" refer to cancer bioassays? Please clarify.

Page 11, 2nd full paragraph – This entire paragraph appears very statistically-oriented and almost seems to abandon the science. For the subject matter at hand (i.e., development of interim human health benchmarks), this discussion is too abstract and makes the reader wonder whether statisticians were the primary authors of this section. Please make a case for why there needs to be so much attention and emphasis on statistical analysis and theory.

Page 12, 3rd full paragraph – Same comment as above. Far too statistical in nature. Mathematics and statistics do not govern carcinogenesis or the propensity for a chemical to act as a carcinogen. Minimize this particular discussion.

Page 12, number 2. – The R value of 0.15 seems very conservative. Can an example be shown that demonstrates how this R value is applied and how it alters the final CSF? Please provide an example to show by how much it reduces the final interim benchmark. By two-fold, ten-fold, or greater?

Page 12 – There is some concern about using an R value that has been derived from 70 different constituents and whether there is enough chemical specificity embedded within this value so as to have any meaning when it is applied to specific interim benchmarks. In other words, the R value appears to be an averaged value based on many different constituents. Perhaps this is the intent and if so, there should be some mention that it is a generic (non chemical specific) value.

Page 13, 1st paragraph – "…will result in human health benchmarks that are overprotective…." " The degree of overprotection is not known, using these methods." This reviewer is concerned that there may be so much protection, or conservatism associated with this approach, that exit levels may be well below background concentrations, and thus meaningless from a risk assessment perspective. If an interim methodology sets a level that is below natural or background levels, then what is the utility of this approach in terms of its true measure as a health protective approach? Again, where has the biological meaning and relevance gone? There is again too much emphasis on statistical analysis in this approach.

Page 13, Boxed area – When a sulfate or hydrochloride form of a constituent is used as a surrogate, it certainly should be chemical specific. There are many such forms that are much different in toxicity than the parent compound.

Page 15, 3rd paragraph – Insert "a" into "….may be performed to produce a provisional…."

Page 15, 3rd line from bottom – Hyphenate "chemical's"

Page 16 – Small detail, but insert "CPDB" after the title is first introduced, as later on it is used standing alone, and the reader has to again determine what it is an acronym for.

Page 17, Section 3.1.3.2 – This brief discussion serves to point out well, how much statistical analysis and emphasis is embedded within the approach. There is simply not enough discussion pertaining to the biology and toxicological properties of various constituents and how these influence a chemical's carcinogenic potential.

Page 21, 2nd bullet – Should this be a "statistically sufficient" number of test subjects?

Page 21, 1st bullet – Unlikely to have a chronic or subchronic study in humans.

Page 21, 3rd paragraph – Excellent point about selecting a NOAEL or LOAEL in the most relevant species.

Page 21, last paragraph – This discussion is all standard knowledge and it is probably not needed here.

Page 22, paragraph beginning with "Uncertainty factors of 1,3……" Again, this is standard and common knowledge and probably not necessary within this document.

Page 22 – "…there is no way of knowing how much confidence should be placed in the claim that any true threshold for effect in the most sensitive human subpopulation will be below the RfD or RfC value." It should be noted that thresholds may be above the RfD or RfC and there is the possibility that they are far above these health benchmarks.

Page 23, last line – "….more heterogeneous…." In what respect? With respect to genetics, metabolism, hair color, etc. Please specify.

Page 24 – "Pseudo-GSDs, that are the square root of the 84th percentile divided by the 16th percentile, were provided by the study authors because the actual distribution is not lognormal but is characterized non-parametrically." This reviewer has serious doubts as to whether many scientists or risk assessors will have any idea as to what this sentence means. This is a prime example where the approach and documentation contained with this Draft, is too statistical in nature. What does this sentence have to do with toxicity, biological effects, biological meaning, etc.

Page 25 – "Application of the adjustment factors presented by Baird et al. (1996) however, would necessarily produce an interim noncancer benchmark that is less conservative than the standard approach, violating a primary goal of the expedited approach." Yet, what if application of the factors proposed by Baird are more meaningful or accurate from a scientific perspective, ones that yield a value that is closer to the true mark. The policy of upholding conservatism in this approach could not be spelled out more clearly than in this paragraph and this makes the reader question whether the OSW is more interested in identifying and developing accurate health benchmarks or in insuring that any developed benchmark is more conservative than would be traditionally derived.

Page 26 – "…would result in interim noncancer benchmarks less conservative than the standard approach, which would be inconsistent with a primary goal of the expedited approach." Who cares whether or not the approach is less conservative, what we should be interested in is biological meaning, scientific accuracy, and in developing a benchmark that is grounded in science and that approximates the best estimate of a true health benchmark. The policy discussion should be placed elsewhere. The authors appear to be discussing various approaches, but then discounting their value and utility because they violate the "conservative nature" that is desired within this approach.

Page 26, 2nd to last line – "select a benchmark from the distribution of existing benchmarks" Benchmarks for what effects and for what chemicals?

Page 27 – "In general, the toxicities of salt, sulfate, and acetate compounds are not expected to differ significantly from their parent compounds." Be very careful with this statement. There are many such compounds whose toxicity varies significantly from the parent compounds and this is an over generalization.

Page 30 – "to an RfD of 0.00008 mg/kg/day)." This is a relatively low RfD and one wonders whether through statistical application of the log function and the resulting low value if there is any biological meaning in this RfD.

Page 30 "The advantage of this approach is that it is simple, quick, and requires no chemical-specific information." This is what is worrisome in that there is absolutely no consideration of the biology, chemistry, pharmacokinetics, toxicology, or any other parameter with more

relevance that simply statistical manipulation of RfD or RfC distributions.

Page 30 "The interim exit concentrations, however, are likely to be significantly lower than those that would be developed under a traditional approach, and the degree to which they are lower cannot be determined." This is unacceptable as an explanation and the authors must anticipate or forecast what will develop if exit concentrations or interim benchmark values are established that are lower than background, natural, or even detectable concentrations. This serves no purpose and if such is the case, then an interim approach should be abandoned.

Page 35, last paragraph – Better to use ranges of health benchmarks within chemical classes than to use the entire distributions.

Page 38 "Expedited Review" – Obviously, while the authors admittedly are desiring input on this section and perhaps that is why there is a general lack of critical information on the review process, there needs to be some further development of the specifics associated with this part of the process. It is too brief, yet one of the most critical steps in health benchmark development. What is the process that will be proposed? Will external (from the Agency) reviewers be used?

Page 38 – "An expedited review may involve a limited number of reviewers,…. Relaxing the need for consensus, and allowing for greater uncertainty so long as the interim health benchmarks are likely to be lower than those that….." Is this what we want out of such an approach, for their to be no need for consensus? For there to be greater uncertainty, so long as we maintain the conservatism of the developed values? If there are a limited number of reviewers, there had better be a diversity of disciplines represented, particularly risk assessors, toxicologists, and those versed in safety assessment.

Page 38 – "Use of the expedited approach to develop interim cancer and noncancer benchmarks will result in values more conservative than would use of a standard approach." Yet, in actuality, any approach could be used as long as one simply increased the uncertainty factor sufficiently. What makes this approach unique or different from the standard approach that would increase the UF in developing health benchmarks?

Page 38 – "It should be noted that interim health benchmarks may change as more information becomes available……" The Agency must have in place some mechanism or opportunity for external scientists to provide additional information and data that may be useful in changing an interim value. What are the provisions if an external scientist wanted to submit an alternative value based on sound and defensible data?

Page 39 – Change "in this report should be applied" to "could be applied"

Page 39 – Same change in two places "following methodology could be used" and "the following methodology could be used"

Page 40 – Top two lines – This is redundant. Authors have already detailed these points. Omit.

Page 40 – "relaxing the need for consensus" This is not necessarily good.

Page 40 – 1st paragraph – All the same points as previously mentioned: Who will the reviewers be, is it necessarily good to have greater uncertainty, etc.

Page 40 – Last paragraph – This is a poor finale or summary of the entire document and the authors should attempt to bring the focus together at this point. Why end the entire document with a sentence on methods available to extrapolate RfCs to RfDs and from inhalation to ingestion cancer benchmarks?

Appendix A-1 – Is the TD50 geometric mean a mean of both rat and mouse TD50s? In other words, have TD50s from different rodent species been averaged and if so, do the authors consider this a scientifically defensible approach?

ATTACHMENT C

Comments on*"Conceptual Approach to Establish Interim Human Health Benchmarks"*

Gary Pascoe

DABT

Review of

Conceptual Approach to Establish Interim Human Health Benchmarks

Research Triangle Institute, 1999

Prepared by

Gary A. Pascoe, Ph.D., DABT

for

Eastern Research Group, Inc.

Lexington, MA

11 November 1999

This review follows the format of the Issues for Review in the *Charge to Reviewers*, and includes specific comments on sections of the report

General Issues for Review

1. As pointed out in the report, alternative methods that are considered less rigorous than standard acceptable methods will be unacceptable to a portion of the scientific community. As that may be, the overall concept of deriving interim benchmarks and the methods proposed in the report are defensible within the context of their intended use and purpose. I believe that one of the larger problems with the overall concept is not so much the technical approach to deriving the benchmarks, but the subsequent use of the benchmarks in applications they are not intended for. It is common for both regulators and the regulated community to rely on rapid means or readily available numerical benchmarks for assessing risks from public exposure to chemicals. This has often been the case in evaluating contaminated properties and other uses of risk assessment outside of federal agency oversight. Any final product or benchmark developed under the proposed expedited approach must have cautionary language against the application of the benchmarks to unintended uses, such as cleanup of hazardous waste sites.

The major sources cited in the report are from highly regarded peer review journals and are sufficient for the intended purposes of the approach. I am not aware of other sources that have undergone a peer-review

process.

2. Of alternative approaches to developing interim benchmarks that I have encountered, none are recommended for inclusion in the report. I am not aware of alternative approaches that are more appropriate or more defensible than those presently in the report.

3. In general, I agree with the logic of the approach hierarchy as illustrated in Figure 3-1; however, see my comments on the use of TD50s for deriving CSFs. The multistage cancer dose model should be given priority, but following an initial identification of a carcinogen from the database of TD50s.

4. The approach described for using chemical analogs where dose-response data are lacking is a conservative approach. On the surface, the limitations that are planned appear to be more restrictive than necessary. However, there are no additional chemical groups to which a QSAR-type approach could be applied in a clearly defensible manner. Data are insufficient to comfortably determine what other chemical groups or moieties could be used in a relationship with toxicity. As such, the approach as planned should be considered sufficient.

5. A permanent, and paid, external review panel should provide a faster avenue for review than using internal agency personnel.

Cancer Issues for Review

1. I think that the use of an uncertainty factor, with the linearized multistage model, based on the ratios of cancer benchmarks developed by both approaches is very appropriate, and adds just enough conservatism to justify the expedited purpose of the benchmarks.

2. An argument that the use of the linearized multistage model will be too difficult or time-consuming to be appropriate to an expedited benchmark development process has not been sufficiently developed in the report to rule it out as an approach. An estimated effort for using the cancer model and using the TD50s should have been better explored in order to provide more justification for the TD50 method. Because the regression for TD50 and CSF values is so strong, it should be retained as a next tier method if the dose-response data are insufficient for use in the multistage model.

3. Section 3.1.3 points out that the determination as to whether a chemical is carcinogenic could be time consuming using the multistage model approach. Since the CPDB can be used to identify carcinogenicity, why not use it for that step of the process, then use the cancer dose model to calculate the expedited CSF? And then include the R factor for a measure of conservativeness and uncertainty in the expedited approach.

Non-Cancer Issues for Review

1. Although the Evans & Baird and the Swartout approaches provide a way to deal with uncertainty in the RfD estimates, applying them to an expedited approach with limited data seems inappropriate. Primarily because they result in less conservative RfDs than using a traditional approach. The summary text on page 26 points out the problem with using either approach because of the lack of conservativeness that should be inherent in an expedited approach. However, the Swartout approach still appears to be useful, particularly in the case of limited data, which is probably likely. Why not use the 99 percentile UFs of Table 3-2? These would result in UFs slightly more conservative than traditional approaches, and would retain some of the benefit of an understanding of the confidence in the RfD values.

The discussion on page 26 about the Evans & Baird and the Swartout approaches appropriately points out the advantages of better accounting for the confidence in the RfDs. It may be useful to also point out that there is increasing approval for applying these approaches to the IRIS database to more accurately calculate RfDs. Once that is done, the application of these approaches at a more conservative level to the RfD derivation would be appropriate for an expedited approach. Prior to that happening, the use of the approaches is pre-mature, largely because of the problem of resulting in less conservative RfD values, unless the 99 percentile UFs are used in the Swartout approach, as suggested above.

2. I believe that the approach to using TLVs to approximate RfCs is the best one currently available for an expedited approach, and is a popular approach frequently suggested to regulatory agencies. It should be pointed out that the approach is appropriately conservative by using the lower confidence limit, especially given the intent of the TLV to protect workers exposed under different conditions than the public, but that it could result in an RfC up to 3 orders of magnitude different than a value determined by traditional methods.

3. I support the use of the proposed distribution method for the intended purposes of an expedited derivation of a benchmark, especially given a lack of alternatives that offer more confidence than the proposed method. In this regards, the cautionary statement mentioned under General Issues above about the likely use of the interim values is important.

4. The use of the SAR scheme to the distributional analysis of RfCs and RfDs is definitely an improvement to the approach, and removes an initial concern about the large range of values that would otherwise result. No other chemical classes would improve the analysis, mostly because of the low number of chemicals that could possibly be placed in any further breakdown of classes.

Specific Comments

Section 3.2.1. This section is a little difficult to follow at first. It would help greatly to have two subheadings: "Traditional Process" at the start of the first paragraph, and "Expedited Approach" at the start of the paragraph immediately following the equation for the RfD on page 22.

It is not clearly stated, but appears that the expedited approach would follow the traditional approach described in the first set of bullets under Section 3.2.1, page 21. This needs to be stated at the beginning of the "Expedited Approach" section on page 22, followed by the discussion of the expedited approach to the uncertainty factors.

Page 24, Accounting for Incomplete Database – The last sentence of the first paragraph is a little confusing. It would help to rearrange as "Evans and Baird examined…and calculated how much lower the missing NOAEL might be for each…"

Page 24 - The equation is unclear as to what the "Available minimum NOAEL" refers to. Is this a ratio of the NOAEL from each study to the minimum NOAEL in the database?

Page 24 – A tie-in between the data in the Evans and Baird paper on accounting for a missing NOAEL and the issue of accounting for an incomplete database would be helpful at the end of the Evans and Baird discussion.

Section 3.2.2.1, second paragraph, last sentence – The use of pyrene as a surrogate is repeated at the beginning and end of the sentence.

Section 3.2.2.2 – It is not clear how adjustments to the benchmark would be made to account for differences in pharmacokinetics, etc. Some citations to literature that discusses this would be useful here.

Section 3.2.2.3, first paragraph, last sentence – Suggest adding to the end "…extended period, *and often the values are based on the same toxicological or epidemiological studies*."

Section 3.2.2.3, second paragraph – The logic of the sentences and bullets is unclear. Suggest dropping the bullets and change the text to "…are provided in Appendix B. *Based on the regression relationship, a procedure for developing an RfC from a TLV is to* determine the TLV for the constituent of interest and select…"

ATTACHMENT D

Comments on"*Conceptual Approach to Establish Interim Human Health Benchmarks*"

Lauren Zeise

California Environmental Protection Agency

Review:

A Conceptual Approach to Establish

Interim Human Health Benchmarks

The overall strategy to establish benchmarks through an expedited strategy is a good one and the Agency has been thoughtful in its approach to the matter. However some of the techniques outlined required adjustments to reduce the uncertainty and bias.

Charge Questions

General Questions

1. Do you feel the major sources cited (e.g., Hoover et al., 1995) have gained sufficient acceptance in the risk assessment community for application in a regulatory context?

Expedited Cancer Potency Development

Values generated by the expedited technique of Hoover et al. (1995) are being used in various regulatory programs in state government risk assessments in California. Expedited values have been employed in air toxics "hot spots" assessments, hazardous waste evaluations, and Proposition 65 enforcement. Thus they have achieved regulatory acceptance. It is important to acknowledge that different risk assessment needs require different levels of evidence, and that the expedited values are primarily being used in "right to know" context (Proposition 65 and Toxic Hot Spots). When the regulatory impacts are more far reaching, such as widespread regulation (e.g., of an air toxic contaminant), more labor intensive methods are employed. Also, risk assessment structure for establishing potencies in California under programs utilizing expedited values is hierarchical, with differing levels of review and appeal, depending on the level of scientific investigation involved in establishing the number (see e.g., California Code of Regulations, Section 12705).

It appears that the risk assessment context under consideration for the application of an expedited approach would benefit from the use of such an approach. That being said, there are important technical differences between the approach described in the document and that of Hoover et al., so answering the charge question in the affirmative can not be seen as an indication that the approach laid out in the OSW document should be used in a regulatory context.

2. Are there alternative methods not presented in this report that you would recommend for using in deriving interim health benchmarks?

There are other toxicity based approaches that could be considered (e.g., Zeise et al., 1984; Travis et al.). There will be less general consensus about the use of these approaches, so they should be employed only if the Agency believes it can withstand the criticism for doing so. Such approaches would enable the coverage of a greater number of chemicals with carcinogenic activity than is provided for in the current scheme.

3. Do you support the logic illustrated in Figure 3-1 hierarchy for development of interim benchmarks? If not, what changes would you recommend to strengthen the hierarchy?

Selection of Chemicals for Consideration

The figure gives no indication of how chemicals will be selected to enter the identification scheme provided in the figure. Of the myriad of chemicals of potential concern, how will the Agency decide which ones to focus on? Will it be done on the basis of availability of toxicity information? Structure activity? What quality assurance will be in place to ensure that those of greatest risk potential are being addressed first? Some discussion of this issue would be helpful, and if possible, it should be addressed in Figure 3-1.

Carcinogen Identification

The first step in Figure 3-1, how agents are identified as a carcinogen or not, should be reconsidered. The main problem with this approach is the limited number of chemicals that can be addressed by it. The approach is being applied to a wide variety of agents, many of

which may not have been adequately tested for carcinogenic activity or undergone the labor intensive problem of carcinogen classification by the named authoritative bodies. The number of chemicals which have been adequately tested in animal bioassays and reviewed formally as part of carcinogen classification may be a very small fraction of the number of chemicals potentially of concern to OSW. Thus, the process would benefit from an expedited hazard identification component. While a comprehensive approach to this problem would be hard to develop, receive regulatory acceptance and implement in the near term, some relatively modest improvements could be made and implemented under the current scheme. Meanwhile, the Agency should, in a long term effort, work on developing a more comprehensive approach. In this regard collaboration with scientists in the Agency's premanufacturing notice program, and consideration of approaches employed in the pharmaceutical and pesticide product development should be beneficial.

The initial carcinogen identification question asked should not be whether or not some body has identified an agent as a carcinogen but the extent to which the agent may pose a carcinogenic hazard. Findings of authoritative bodies such as IARC, EPA IRIS and National Toxicology Program (NTP within DHHS) will help answer this question, but if the chemical has not been reviewed or has been reviewed some time ago, its carcinogenic potential should still be assessed. First, at the simplest level, if two organic salts dissociate into the same active ionic form, and one is listed by one of the three authoritative bodies but the second is not, the Agency should treat the second as a carcinogen. Second chemicals that are metabolized to known/likely carcinogens should be treated as carcinogens. Third, and more difficult, if there are clear structure activity relationships *and* the chemical fits within a class of chemicals with known carcinogenic activity (e.g., nitrosamine, nitrosourea) then it may be appropriate to include it in the group of chemicals with likely carcinogenic activity. Fourth, if a chemical has been studied in bioassay and clearly tests clearly positive, or has strong positive mechanistic information and suggestive information from the bioassay, it may also be appropriate to treat it as a carcinogen. Fifth, if an authoritative institution besides the three named (e.g., National Institute of Occupational Safety and Health) has identified the chemical as possessing carcinogenic activity or the chemical is on California's Proposition 65 list, it could be treated as such. The existing scheme could be expanded to include these approaches.

Finally, one could investigate the application of a more general approach for identifying carcinogens based upon structure activity and other considerations, but this would take time to develop and could not be included in the scheme for some time. It nonetheless remains problematic to identify chemicals as not possessing carcinogenic activity because of paucity of information. Practical approaches may be developed that are not as data intensive as those of the current classification schemes used by the Agency, IARC and NTP. Those employed in drug and pesticide development could be considered and expanded upon. Ultimately it may be beneficial at some future point to move a probablistic scheme.

A second problem with the scheme proposed is that it may be uneven in carcinogenicity classification. The Executive Summary indicates that EPA group C agents will be treated as carcinogens, but for IARC, those in groups 1 and 2 would be included. Group 3 IARC evaluations finding limited evidence in animals are roughly equivalent to EPA's group C classifications. It seems appropriate to include chemicals with limited evidence as possessing some carcinogenic activity, but some adjustments may be warranted in quantifying their activity. In the case where a chemical has been tested to a limited extent (e.g., limited pathological evaluation, single study) but positive finding resulted and where mechanistic studies suggest a genotoxic mechanism, the application of the approach could be the same as for chemicals with more evidence (e.g., IARC 2B). In the case where the chemical has been extensively tested and found to have weak activity in a few test, and mechanistic information is not impressive, a less stringent approach may be in order. For this case, it may be appropriate in certain cases to apply a standard non-cancer approach to the bioassay dose response data and include an extra uncertainty factor. A benchmark dose would be preferable to a NOAEL as the point of departure for this case. Non-cancer endpoints should also be evaluated to determine if they produce a lower benchmark. Alternatively, as is sometimes done by the Agency in its drinking water program, the benchmark may be set on non-cancer findings, with an extra uncertainty factor applied to account for possible carcinogenicity. This is less preferable but has the advantage of being already in use within the Agency.

Cancer dose response evaluation

The scheme does not indicate how it will address agents which are identified as having carcinogenic activity but do not have data sets meeting the selection criteria (e.g., too few dose groups). The document suggests use of an expanded structure activity approach, and exploration of such techniques for this use are encouraged. Another possibility is to include a toxicity based scheme to estimate carcinogenic potency (e.g., Zeise et al., 1984). It may be that some hybrid of these two approaches would lead to better estimates. This should be explored. It would greatly increase the coverage.

A major concern regarding the approach laid out in Figure 3-1 is the inclusion of a separate TD50 carcinogenicity measure. If a TD50 exists in the Gold et al. database then dose response data, tabulated in that database, are available for cancer dose response (multistage) modeling, and should be considered for use. There is no reason to rely on a TD50 measure generated by Gold et al. simply because their database is being used. A multistage analysis should be applied to those data. This will reduce errors introduced when data sets are non-linear. The expedited approach of Hoover et al. involves using the Gold et al. database as a tool for identifying studies for analysis, and also using the tabulated data as the basis for analysis in most cases. The exception is for cases where

there is high intercurrent mortality among treatment groups. In such a situation an attempt should be made to obtain individual animal data and a time dependent multistage analysis performed.

It is suggested that the same approach be applied here. Selection of a suitable bioassay data set may come from either the Gold et al. database or through literature review but would feed into the same dose response analysis. It would be more expeditious to start with the Gold et al. database and supplement that with additional good studies identified in the literature.

Gold et al. have already applied a series of criteria to the available data; those criteria are not as stringent as the ones laid out in the document. It is noted though that the thorough evaluation of studies with respect to criteria laid out in the document and documentation that such criteria are met will not result in an expedited approach. Also, stringent application of criteria may result in exclusion of valuable datasets, especially if the requirements that the bioassay contain multiple doses (as indicated in the Executive Summary) and peer review are adhered to. Another concern is the strict application of route criteria. Perhaps instead the conceptual approach for cancer dose response analysis should ask -- what is the best approach for characterizing the carcinogenic activity for this chemical? The criteria should be used as a guide for selecting the best studies but should not preclude the use of the dose response data or force a TD50 estimate when they are not met.

Non-cancer dose response evaluation

The cascading scheme for determining non-cancer benchmarks appears reasonable and defensible. Succession of methods used should proceed from those producing the most certain result to those of less certainty. It is not clear that this is necessarily the case in terms of the order of the chemical analog analysis and the TLV/RfC regression. Some further exploration of this issue to evaluate which one results in more certain estimates is needed. Again, too strict application of the criteria for study evaluation may have the unintended result of moving one toward a less certain analysis and result. As for cancer endpoints, the criteria laid out for study selection should be characterized as study selection guidance. Some minimal data requirements for dose response analysis will be needed, but they should be far less stringent than those suggested by the criteria.

4. Do you believe that, given the goal of the expedited methodology, these

approaches are scientifically defensible or should be expanded to include additional techniques and constituent groups?

As discussed above, the Gold et al. database is a valuable source of compiled dose response data to which a multistage analysis should be applied. In cases of high intercurrent mortality time dependent analyses should be performed if data are available. In application of the approach standardized rules for correction of study length (assuming say risk increases with the third power of age) and adjusting for route of exposure should be employed.

Stringent application of criteria for study selection is not scientifically defensible. The resulting exclusion of studies will push the evaluation toward less certain techniques. For example, the document states that for inhalation benchmarks only inhalation or intratracheal routes of exposure will be used. For the case for which there is only an oral bioassay for cancer and an inhalation number is needed, consideration given to route adjustment. Also stringent application of criteria may result in exclusion of valuable datasets, especially if the requirements that the bioassay contain multiple doses (as indicated in the Executive Summary) and peer review are adhered to. The criteria instead should be used as guidance, with the overall objective of obtaining the most defensible estimate of potency given the data sets and techniques available for any particular case.

Also, the expedited methodology should include a component for expedited cancer hazard identification, also discussed above. Use of information on different organic salts which dissociate to the same active form is appropriate, and the same logic should also apply at the hazard identification phase. Care should be taken in evaluating the quantitative activity of metallic salts; mischaracterizations can lead to significant misestimations of cancer potency. The use of TEFs in the context of the OSW program is encouraged as is the expansion of TEFs beyond dioxins and PAHs to other classes of chemicals. Research to establish basic relationships on which to establish TEFs for other classes of chemicals is encouraged. Further expansion of the expedited approach to include toxicity based schemes for those chemicals for which bioassay data are limited or non-existent is also encouraged. Greater exploration of structure activity relationships for use in expedited cancer potency assessment is also encouraged. Such approaches might be more accurate if integrated with toxicity based approaches. More work in this area is clearly needed.

5. Recommendations for expedited review

Clearly the development of expedited values should not be subjected to the same level of peer review and external scrutiny as consensus IRIS values. The document should provide more definition to the review than given in the report. It is recommended that internal review of values be provided by the National Center for Environmental Assessment NCEA on an expedited basis, and that they also undergo limited external review.

Cancer Issues

1. Do you agree with the proposed use of the R factor to account for uncertainty inherent in the expedited approach? If not, please recommend alternatives for consideration by OSW.

The proposal does not follow the expedited approach of Hoover et al. and the R factor does not represent an appropriate adjustment. Please note that overall the expedited method discussed by Hoover et al. resulted on average in more public health conservative estimates than the traditional approach. This is in part due to standardization of the practice of potency estimation subsequent to the development of some of the potency values used by the Agency. The quantitation and data selection procedures for the expedited approach should mirror as much as possible the current traditional approach. Thus it would be preferable to rely on literature reviews and the Gold et al. database for the identification of the benchmark study and to apply a standard modeling procedure to the set. This would include a standardized approach for adjusting for study length, lack of fit of dose response relationships, dose calculations, and interspecies adjustments. No additional factor is needed.

2. Do you agree with the use of the relationship between TD50's and CSFs through regression analysis to derive interim potency values? If not, can you suggest other methods that you consider appropriate for this application?

The TD50-CSF relationship should not be used. Rather the multistage model should be fit to the dataset used to derive the TD50 values and standard adjustments for study length, lack of fit of dose response relationships, dose calculation, and interspecies differences in body size made.

3. The use of the TD/CSF relationship provides a much quicker, less resource

intensive method of cancer benchmark derivation than using the linearized multistage model. Do you think that these advantages are sufficient to justify a change in the methodology such that this method is used in lieu of the linearized multistage model for constituents that have quantitative cancer data available?

I disagree entirely with the characterization of the application of the linearized multistage model as resource intensive. Results using readily available software packages can be produced within a matter of minutes. The use of the linearized multistage procedure (including the approach taken for poor data fits) is clearly preferable and more scientifically defensible because non-linearities will be better addressed. Results will be more consistent with the traditional approach.

Noncancer Issues

1. Do you believe that the use of probability distributions is an appropriate alternative to the traditional uncertainty factors given the limitations of the supporting data set? If not, would you recommend specific changes to how the probability distributions are used or can you suggest an alternative approach that is more technically defensible?

Although in general the approach of using probability uncertainty factors as laid out by Baird and colleagues may be preferable to the use of the default approach for noncancer endpoints, it does not appear to explicitly address the uncertainties associated with the expedited approach. Potentially the result may be considerably less conservative values than the traditional approach. Although the method shows promise, it is recommended that it undergo further evaluation, particularly within the context of expedited assessment before it is employed by OSW for this purpose. Of particular concern is the lack of data on sensitive human subpopulations (lifetime exposure) for establishing such distributions as well as the lack of understanding of age dependent differences in sensitivity. It is inappropriate to use animal data for derivation of this factor as was apparently done by Baird because of the homogeneity in research animals.

2. Do you consider use of the TLV-RfC regression appropriate and can you suggest improvements to the approach?

The approach appears to be reasonable, especially given the use of lower confidence limit for predicting the RfC. Please note that the uncertainty in using this procedure may be less than that for the chemical analog. The magnitude of the uncertainties associated with the two different process should be more fully evaluated before settling on which one is preferable in the derivation hierarchy (e.g., Figure 3-1).

3. Do you agree that the use of representative distributions is appropriate for the establishment?

4. Estimating interim benchmarks as a function of class.

I agree with the use of distributions in cases where data are absent. It is similar to the concept of toxicological insignificance applied to food additives. It would be preferable to augment the approach with chemical class information. The approach laid out in the document for doing this appears reasonable.