

Compilation of Individual Comments from Panel members

(as of October 12, 2010)

DR. HARVEY CLEWELL	2
DR. TONY COX	11
DR. ELAINE FAUSTMAN	18
DR. SCOTT FERSON	20
DR. JEFFREY FISHER	36
DR. HELEN HÅKANSSON	39
DR. RUSS HAUSER	42
DR. B. PAIGE LAWRENCE	46
DR. MICHAEL LUSTER	50
DR. PAOLO MOCARELLI	57
DR. VICTORIA PERSKY	58
DR. SANDRA PETERSEN	61
DR. KARL ROZMAN	70
DR. ARNOLD SCHECTER	79
DR. ALLEN SILVERSTONE	86
DR. MITCHELL SMALL	89
DR. ANNE SWEENEY	98
DR. MARY WALKER	104

Dr. Harvey Clewell

Preliminary Comments – July 12, 2010

Section 3. The Use of Toxicokinetics in the Dose-Response Modeling for Cancer and Noncancer Endpoints.

3.1 The 2003 Reassessment utilized first-order body burden as the dose metric. In the draft Response to Comments document, EPA used a physiologically-based pharmacokinetic (PBPK) model (Emond et al., 2004, 2005, 2006) with whole blood concentration as the dose metric rather than first-order body burden. This PBPK model was chosen, in part, because it includes a biological description of the dose-dependent elimination rate of TCDD. EPA made specific modifications to the published model based on more recent data. Although lipid-adjusted serum concentrations (LASC) for TCDD are commonly used as a dose metric in the literature, EPA chose whole blood TCDD concentrations as the relevant dose metric because serum and serum lipid are not true compartments in the Emond PBPK models (LASC is a side calculation proportional to blood concentration).

Please comment on:

3.1.a. The justification of applying a PBPK model with whole blood TCDD concentration as a surrogate for tissue TCDD exposure in lieu of using first-order body burden for the dose-response assessment of TCDD.

The use of body burden in the 2003 Reassessment represented an improvement over the usual default metric of administered dose (mg/kg/d), because the default metric would not properly reflect the accumulation of dioxin in the tissues over time. However, because the accumulation of dioxin in fat vs. liver is dose-dependent, body burden would not serve as a direct surrogate for tissue exposure. The use of whole blood concentration in the Response to Comments document is a better choice, because it is more directly related to the most biologically relevant toxicokinetic metric, which is the free concentration of dioxin in the target tissues (liver, fetus, etc.). Blood concentrations are routinely used to estimate biologically effective exposures for pharmaceuticals.

I do not, however, agree with justifying the use of whole blood concentration rather than lipid adjusted serum concentration (LASC) just because that is not the way the Emond model was structured. It would be trivial to change the model so that LASC could be used. The question that should be addressed is only whether whole blood concentrations or LASCs provide better surrogates for cross-species and cross-study comparisons of free dioxin concentration in the target tissues. LASC is the preferred measure for reporting dioxin biomonitoring data, and is the measurement reported in most of the human epidemiological studies. LASC is also more likely to reflect free

dioxin concentration in the plasma, and hence free concentration in the target tissue. The EPA points out (p. xxxiv) that the LASC is related to the whole blood concentration by a scalar; however, they incorrectly conclude that the metrics are equivalent. In fact, they later (p. 3-511, line 6) discuss the fact that the relationship between them is subject to inter-individual and inter-species variation. They even estimate a 4-fold variation in the relationship of whole blood concentration to LASC in the mouse (p. 3-55, line 13). It's not clear to me at this point how well this issue is addressed in all of the dose metric calculations, but the EPA does use the Emond model to estimate LASCs in the RfD calculations (e.g., p. xli, line 21). Consideration of this issue is unlikely to significantly affect the outcome of the risk calculations, but it would be important for a quantitative uncertainty analysis.

3.1.b. The scientific justification for using the Emond et al. model as opposed to other available TCDD kinetic models.

I am satisfied with the justification for the use of the modified Emond model for the dose metric calculations in the assessment.

3.1.c. The modifications implemented by EPA to the published Emond et al. model.

The EPA modifications are minor and appear to be appropriate.

3.1.d. Whether EPA adequately characterized the uncertainty in the kinetic models.

The EPA document presents a reasonably thorough qualitative characterization of the uncertainty in the kinetic models, sufficient to support their use in the assessment. A more quantitative uncertainty analysis would be beneficial in the risk characterization, if only to demonstrate the relatively low contribution of the kinetic modeling to the overall uncertainty in the assessment.

3.2. Several of the critical studies for both noncancer and cancer dose-response assessment were conducted in mice. A mouse PBPK model was developed from an existing rat model in order to estimate TCDD concentrations in mouse tissues, including whole blood.

Please comment on:

3.2.a. The scientific rationale for the development of EPA's mouse model based on the published rat model (Emond et al., 2004, 2005, 2006).

I am satisfied with the approach used to develop a mouse model on the basis of the published rat model and the available mouse kinetic data.

3.2.b. The performance of the mouse model in reference to the available data.

The mouse model performs very well, and is clearly adequate for use in estimating dose metrics for the assessment.

3.2.c. Whether EPA adequately characterized the uncertainty in the mouse and rat kinetic models. Please comment specifically on the scientific justification of the kinetic extrapolation factor from rodents to humans.

The EPA provides an adequate characterization of the uncertainty in the mouse and rat kinetic models, sufficient to justify their use, together with the human model, to estimate rodent-to-human extrapolation factors.

3.3 Please comment on the use of Emond et al. PBPK model to estimate human intakes based on internal exposure measures.

The modified Emond model is the best available approach for estimating exposures on the basis of internal exposure measurements. Nevertheless, there is considerable uncertainty associated with attempting to reconstruct prior exposures in a human population (e.g., Serveso).

3.4 Please comment on the sensitivity analysis of the kinetic modeling (see Section 3.3.5).

The sensitivity analysis published by Emond et al. 2006 is entirely adequate.

3.5 Both EPA's noncancer and cancer dose-response assessments are based on a lifetime average daily dose. Did EPA appropriately estimate lifetime average daily dose? If not, please suggest alternative approaches that could be readily developed based on existing data.

I agree with the average daily dose calculation approaches described in the EPA document.

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment

6.1 Please comment on the discussion in this Section. Is the response clearly presented and scientifically justified?

It's difficult to answer this question at the level of the entire section 6. Section 6.1 provides a rather pedantic discussion of terminology that would probably only be clear to another statistician. It might need to be revised from the viewpoint of what needs to be explained to non-statisticians. Section 6.4 is a long, depressing litany of all the reasons why the EPA cannot conduct the quantitative uncertainty analysis that NAS has requested. I wish that EPA had spent half as much time trying to implement even a rudimentary quantitative uncertainty analysis.

6.2 Please comment on EPA's overall conclusion that a comprehensive quantitative uncertainty analysis is not feasible.

What EPA seems to have done is set an extremely high standard for a comprehensive quantitative uncertainty analysis and then explain why it is not feasible. What they do not do is try to actually present any kind of quantitative uncertainty analysis as requested by NAS. There are many approaches that could be used to provide insight into the dispersion of risk estimates from different models/assumptions/decisions in the spirit of the ill-fated OMB risk assessment guidelines. This has been, and continues to be, the key deficiency in the EPA risk assessment for dioxin. This problem is primarily an issue for the cancer risk assessment, where linear and nonlinear options are discussed, the nonlinear options are rejected, and only low-dose-linear results are carried forward.

6.2.a. Please comment on the discussion in Section 6 regarding volitional uncertainty and how this type of uncertainty limits the ability to conduct a quantitative uncertainty analysis.

The quantitative implications of volitional uncertainties can be estimated or at least portrayed using a number of techniques from decision analysis (Clewell, H.J., Andersen, H.J., and Blaauboer, B.J. 2008. On the incorporation of chemical-specific information in risk assessment. *Toxicology Letters*, 180: 100-109.).

6.3 Throughout the document (including the Appendices), EPA presents a number of limited sensitivity analyses (e.g., toxicokinetic modeling, RfD ranges, cancer OSF ranges, cancer RfD development). Please comment on the approaches used, and the utility of these sensitivity analyses in clarifying potential significant uncertainties.

These individual analyses are all worthwhile but they do not take the place of an overall uncertainty analysis during the risk characterization.

Clewell Comments - Updated August 27, 2010

Section 3. The Use of Toxicokinetics in the Dose-Response Modeling for Cancer and Noncancer Endpoints.

3.1 The 2003 Reassessment utilized first-order body burden as the dose metric. In the draft Response to Comments document, EPA used a physiologically-based pharmacokinetic (PBPK) model (Emond et al., 2004, 2005, 2006) with whole blood concentration as the dose metric rather than first-order body burden. This PBPK model was chosen, in part, because it includes a biological description of the dose-dependent elimination rate of TCDD. EPA made specific modifications to the published model based on more recent data. Although lipid-adjusted serum concentrations (LASC) for TCDD are commonly used as a dose metric in the literature, EPA chose

whole blood TCDD concentrations as the relevant dose metric because serum and serum lipid are not true compartments in the Emond PBPK models (LASC is a side calculation proportional to blood concentration).

Please comment on:

3.1.a. The justification of applying a PBPK model with whole blood TCDD concentration as a surrogate for tissue TCDD exposure in lieu of using first-order body burden for the dose-response assessment of TCDD.

The use of body burden in the 2003 Reassessment represented an improvement over the usual default metric of administered dose (mg/kg/d), because the default metric would not properly reflect the accumulation of dioxin in the tissues over time. However, because the accumulation of dioxin in fat vs. liver is dose-dependent, body burden would not serve as a direct surrogate for tissue exposure. The use of whole blood concentration in the Response to Comments document is a better choice, because it is more directly related to the most biologically relevant toxicokinetic metric, which is the free concentration of dioxin in the target tissues (liver, fetus, etc.). Blood concentrations are routinely used to estimate biologically effective exposures for pharmaceuticals.

I do not, however, agree with justifying the use of whole blood concentration rather than lipid adjusted serum concentration (LASC) just because that is not the way the Emond model was structured. It would be trivial to change the model so that LASC could be used. The question that should be addressed is only whether whole blood concentrations or LASCs provide better surrogates for cross-species and cross-study comparisons of free dioxin concentration in the target tissues. LASC is the preferred measure for reporting dioxin biomonitoring data, and is the measurement reported in most of the human epidemiological studies. LASC is also more likely to reflect free dioxin concentration in the plasma, and hence free concentration in the target tissue. The EPA points out (p. xxxiv) that the LASC is related to the whole blood concentration by a scalar; however, they incorrectly conclude that the metrics are equivalent. In fact, they later (p. 3-511, line 6) discuss the fact that the relationship between them is subject to inter-individual and inter-species variation. It's not clear to me at this point how well this issue is addressed in all of the dose metric calculations, but the EPA does use the Emond model to estimate LASCs in the RfD calculations (e.g., p. xli, line 21). Consideration of this issue is unlikely to significantly affect the outcome of the risk calculations, but it would be important for a quantitative uncertainty analysis.

3.1.b. The scientific justification for using the Emond et al. model as opposed to other available TCDD kinetic models.

I am satisfied with the justification for the use of the modified Emond model for the dose metric calculations in the assessment.

3.1.c. The modifications implemented by EPA to the published Emond et al. model.

The EPA modifications are minor and appear to be appropriate.

3.1.d. Whether EPA adequately characterized the uncertainty in the kinetic models.

The EPA document presents a reasonably thorough qualitative characterization of the uncertainty in the kinetic models, sufficient to support their use in the assessment. However, a more quantitative uncertainty analysis is needed for the risk characterization, using Monte Carlo techniques (as in the vinyl chloride IRIS Technical Support Document). It is critical to demonstrate the dependence of human HED and risk predictions on uncertainty and variability in the model parameters, particularly those with high sensitivity (Evans and Andersen, 2000). Moreover, dose metric uncertainty needs to be determined under the same exposure conditions that dose metrics are calculated: both for the various studies that serve as the basis for the dose-response assessments and for human exposures at the corresponding HEDs and risk specific doses.

3.3. Several of the critical studies for both noncancer and cancer dose-response assessment were conducted in mice. A mouse PBPK model was developed from an existing rat model in order to estimate TCDD concentrations in mouse tissues, including whole blood.

Please comment on:

3.3.a. The scientific rationale for the development of EPA's mouse model based on the published rat model (Emond et al., 2004, 2005, 2006).

I am satisfied with the approach used to develop a mouse model on the basis of the published rat model and the available mouse kinetic data.

3.2.b. The performance of the mouse model in reference to the available data.

The mouse model performs very well, and is clearly adequate for use in estimating dose metrics for the assessment.

3.2.c. Whether EPA adequately characterized the uncertainty in the mouse and rat kinetic models. Please comment specifically on the scientific justification of the kinetic extrapolation factor from rodents to humans.

The EPA provides an adequate characterization of the uncertainty in the mouse and rat kinetic models, sufficient to justify their use, together with the human model, to estimate rodent-to-human extrapolation factors. However, a more quantitative uncertainty analysis is needed, using Monte Carlo techniques (as in the vinyl chloride IRIS Technical Support Document) to estimate the propagation of uncertainty from

the PBPK model parameters to the dose metric predictions. On the other hand, formal recalibration of the PBPK model parameters using a Hierarchical Bayesian approach such as Markov chain Monte Carlo analysis is not considered necessary.

3.3 Please comment on the use of Emond et al. PBPK model to estimate human intakes based on internal exposure measures.

The modified Emond model is the best available approach for estimating exposures on the basis of internal exposure measurements. Nevertheless, there is considerable uncertainty associated with attempting to reconstruct prior exposures in a human population (e.g., Serveso).

3.4 Please comment on the sensitivity analysis of the kinetic modeling (see Section 3.3.5).

The sensitivity analysis published by Emond et al. 2006 is not entirely adequate. One of the most important parameters in the PBPK model, the Hill coefficient, is not included in the analysis. Moreover, model sensitivities are species, dose, and dose-scenario dependent, so they need to be determined under the same exposure conditions that dose metrics are calculated: both for the various studies that serve as the basis for the dose-response assessments and for human exposures at the corresponding HEDs and risk specific doses.

3.5 Both EPA's noncancer and cancer dose-response assessments are based on a lifetime average daily dose. Did EPA appropriately estimate lifetime average daily dose? If not, please suggest alternative approaches that could be readily developed based on existing data.

I agree with the average daily dose calculation approaches described in the EPA document.

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment

6.1 Please comment on the discussion in this Section. Is the response clearly presented and scientifically justified?

It's difficult to answer this question at the level of the entire section 6. Section 6.1 provides a rather pedantic discussion of terminology that would probably only be clear to another statistician. It might need to be revised from the viewpoint of what needs to be explained to non-statisticians. Section 6.4 is a long, depressing litany of all the reasons why the EPA cannot conduct the quantitative uncertainty analysis that NAS has requested. I wish that EPA had spent half as much time trying to implement even a rudimentary quantitative uncertainty analysis.

6.2 Please comment on EPA's overall conclusion that a comprehensive quantitative uncertainty analysis is not feasible.

What EPA seems to have done is set an extremely high standard for a comprehensive quantitative uncertainty analysis and then explain why it is not feasible. What they do not do is try to actually present any kind of quantitative uncertainty analysis as requested by NAS. There are many approaches that could be used to provide insight into the dispersion of risk estimates from different models/assumptions/decisions in the spirit of the OMB memorandum: "Updated Principles for Risk Analysis." This has been, and continues to be, the key deficiency in the EPA risk assessment for dioxin. This problem is primarily an issue for the cancer risk assessment, where linear and nonlinear options are discussed, the nonlinear options are rejected, and only low-dose-linear results are carried forward. In fact, both alternatives have support within the scientific community. Both alternatives should be described objectively in the EPA document and the results of analyses based on both alternative modes of action should be carried forward to the risk characterization section of the risk assessment. The EPA has already calculated the necessary points of departure to conduct a nonlinear risk assessment (Table 5-21); they only need to include these results along with the various linear risk estimates (Figure 5-11) as alternatives in the risk characterization to demonstrate the range of possible outcomes for the risk assessment. A figure similar to the characterization of the alternative RfDs in Figure 4-4 could be used. At that point they could still conclude that, based on the EPA guidelines for cancer risk assessment, the residual uncertainty in the mode of action for dioxin carcinogenicity makes it necessary to select the health-protective linear option for low dose extrapolation.

6.2.a. Please comment on the discussion in Section 6 regarding volitional uncertainty and how this type of uncertainty limits the ability to conduct a quantitative uncertainty analysis.

The quantitative implications of volitional uncertainties can be estimated or at least portrayed using a number of techniques from decision analysis (Clewell, H.J., Andersen, H.J., and Blaauboer, B.J. 2008. On the incorporation of chemical-specific information in risk assessment. Toxicology Letters, 180: 100-109.).

6.3 Throughout the document (including the Appendices), EPA presents a number of limited sensitivity analyses (e.g., toxicokinetic modeling, RfD ranges, cancer OSF ranges, cancer RfD development). Please comment on the approaches used, and the utility of these sensitivity analyses in clarifying potential significant uncertainties.

These individual analyses are all worthwhile but they do not take the place of an overall uncertainty analysis during the risk characterization. For example, as mentioned in the comments on Section 3, a quantitative uncertainty analysis of the PBPK model is needed in the risk characterization, using Monte Carlo techniques (as in the vinyl chloride IRIS Technical Support Document). It is critical to demonstrate the

dependence of human HED and risk predictions on uncertainty and variability in the PBPK model parameters, particularly those with high sensitivity, such as the Hill coefficient (Evans and Andersen, 2000). Moreover, dose metric uncertainty needs to be determined under the same exposure conditions that dose metrics are calculated: both for the various studies that serve as the basis for the dose-response assessments and for human exposures at the corresponding HEDs and risk specific doses. The comments provided to the panel by Dr. Tom Starr highlight the fact that alternative values of a single parameter in the PBPK model (the Hill coefficient), both of which values are consistent with the available high-dose animal data, can lead to an order of magnitude difference in risk estimates at the much lower environmental concentrations that are of concern for the risk assessment calculations.

Dr. Tony Cox

Preliminary Comments - July 8, 2010

4. Epidemiology

The discussion of epidemiological evidence for a causal relation between TCDD exposure and increased risk of cancer draws conclusions that seem to be much stronger than the underlying studies justify. For example, the most recent published update from Seveso found that “All cancer incidence did not differ from expectations in any of the contaminated zones” (Pesatori et al., 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2754980/?tool=pmcentrez&report=abstract.) The Hill-type criteria for evidence of “causality” (strength of association, temporality, consistency, gradient, etc., p. 5-6) do not discriminate between causality and false-positive associations due to multiple testing, multiple comparisons, ignored model uncertainty, model specification error, omitted confounders, residual confounding, unmodeled errors-in-explanatory variables, selection of cut points, variable selection and coding, etc. The conclusion that “In summary, EPA finds the available epidemiological information provides strong evidence of an association between TCDD exposure and human cancer that cannot be reasonably attributed to chance or confounding and other types of bias” does not seem well justified by careful analysis and refutation of alternative explanations and sources of spurious associations, such as those just listed. Insofar as most epidemiological studies have been found to be very prone to type 1 errors (e.g., Ottenbacher, 1998, <http://www.ncbi.nlm.nih.gov/pubmed/9554599>) and related sources of bias (e.g., <http://www.ncbi.nlm.nih.gov/pubmed/17186399>), a more careful and critical review of possible non-causal interpretations of the available epidemiological studies is needed.

5.1 Is the weight-of-evidence characterization scientifically justified and clearly described?

The weight of evidence analysis does not address whether TCDD is a human carcinogen at relevant exposure levels, which remains a crucial question for effective risk management, if the dose-response relation contains a threshold (or zero slope of the dose-response curve at the origin, see p. 5-55), or threshold-like nonlinearities (e.g., due to receptor-mediated pharmacodynamics). The decision to use a linear non-threshold model does not appear to me to be well justified by science. The characterization of TCDD as a human carcinogen may be misleading, compared to a more candid discussion of uncertainties (e.g., that it is possible that TCDD is a weak human carcinogen at high doses, but the evidence is far from conclusive, and it is at least as likely that TCDD at current exposure levels is not a human carcinogen).

The public comments I heard in June suggest that this basic uncertainty, about whether TCDD is a human carcinogen (or has adverse effects in humans) at relevant exposure levels, is not being effectively communicated to the public. By this criterion, the weight of evidence characterization is **not** clearly described. A much clearer characterization of uncertainties is probably essential to inform readers, the public, and policy-makers about what the scientific evidence does and does not establish about the carcinogenicity of TCDD.

5.2 Mode of action.

The available data support EPA's conclusions (p. 5-21) that Ahr activation is probably a necessary precursor step for carcinogenic action (and other adverse effects), and that subsequent steps in the causal pathways leading to cancer (if TCDD causes cancer) are uncertain. However, a default assumption of a low-dose linear, nonthreshold dose-response relation is *not* justified by these conclusions. To the contrary, an appropriate dose-response relation for receptor-mediated carcinogenic effects at low doses (with ligand arrivals at receptors being rare events) will typically be sub-linear at low doses (reflecting the sub-linear kinetics for the probability that enough ligands bind simultaneously to trigger a response). This is not made clear in the discussion of p. 5-53, which only considers equilibrium kinetics, rather than rare-event (non-equilibrium) kinetics appropriate for low levels of exposure.

The use of a low-dose nonthreshold modeling approach seems to me to be very questionable, and not consistent with the partial information that is available about the role of Ahr activation. (The assumption that responses are proportional to dose is not well justified by science and does not describe how receptor-mediated responses typically work at low doses.) Thus, while EPA has focused largely on PBPK modeling, I think that better modeling of nonlinear pharmacodynamics (e.g., Simon et al., 2009, www.ncbi.nlm.nih.gov/pubmed/19776211) is important in order to exploit the partial – but extremely useful – knowledge of Ahr-mediated responses now available. Using a low-dose linear non-threshold default for TCDD does not seem consistent with applying relevant biological knowledge to create more realistic risk models and risk estimates.

The discussion of low-dose linearity (cf p. 5-65) does not seem to me to be clear, well-focused, or technically sound. For example, if Ahr is already being activated by non-TCDD exposures, then the incremental risk from TCDD could presumably be zero (in contrast to the Crump argument that incremental contributions are approximately linear). On the other hand, if Ahr is not already being activated by non-TCDD exposures, then the use of rare-event ligand binding kinetics (leading to a sublinear dose-response relation) would seem to be appropriate. The Portier et al., Crump et al., population variability, etc. arguments appealed to in this section do not address the basic need to consider rare-event kinetics. (By rare-event kinetics, I mean the case in which the arrival of ligands at a receptor is well described by a Poisson process, with mean arrivals per unit time proportional to effective dose rate, and the expected number of bound ligands at any time is small compared to the number needed to trigger a response. The probability that a response will be triggered during any time interval is then a sub-linear function of the effective dose rate.)

Overall, it seems to me that current biological knowledge of MoA, while certainly very incomplete (e.g., we do not even know whether TCDD definitely poses a carcinogenic hazard for humans at relevant levels of exposure) justifies a default assumption of zero slope at the origin, with zero incremental risk in the cases where Ahr activation is either already present for other reasons, or is not already present and is not triggered by TCDD exposure. In other words, the default assumption should be a nonlinear, threshold-type model of receptor-mediated effects.

6.2 (Infeasibility of a comprehensive quantitative uncertainty analysis)

Although a completely comprehensive quantitative uncertainty analysis might indeed be too much to expect, now I think that it is both possible and practical to provide readers with much more useful information about uncertainty. A policy maker might reasonably expect the report to provide insight into major uncertainties and questions such as the following:

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- How likely is it that TCDD is not a human carcinogen at current exposure levels? Full discussion of this uncertainty may help to overcome probability neglect and action bias (Patt and Zeckhauser, 2000, <http://www.springerlink.com/content/k47064873365w720/>).
- What is the probability that reducing TCDD exposures would not reduce cancer risk at all, based on recent epidemiological studies and updates such as Pesatori et al., 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2754980/?tool=pmcentrez&report=abstract?
- What is the probability that reducing TCDD exposures would reduce cancer risk by less than 1 excess cancer case per decade (or per year or per century) in the whole US population, under current conditions?
- What is the probability that reducing TCDD exposures would increase cancer risk (e.g., if the dose-response relation is J-shaped or U-shaped)?
- What is the decision-analytic value of information (VoI) from collecting more information on Ahr kinetics and dose-response before making risk management decisions? Although many members of the public believe that it is imprudent and/or morally wrong to delay tighter regulation of TCDD exposures (perhaps reflecting beliefs that TCDD is a potent carcinogen, developmental toxin, etc.) EPA should provide a thorough quantitative decision analysis that makes explicit the current uncertainties and trade-offs and that shows the conditions under which acting now or postponing action are the optimal actions. Without such quantitative analysis, risk management decisions for TCDD will not be adequately informed, and principles other than those of rational decision-making (e.g., the biases discussed in Sunstein and Zeckhauser, 2010, <http://www.hks.harvard.edu/fs/rzeckhau/Sunstein4-6-09.pdf>) may dominate risk management decisions for TCDD. EPA's uncertainty analysis should provide the (decision and management science) scientific basis for improved decision-making. The current decision to, in effect, punt on quantitative uncertainty analysis is not adequate for informing responsible risk management decision and policy-making, and is not justified.

While I agree with EPA that a quantitative uncertainty analysis is challenging, I do not think that it is impractical to undertake one. It may well be true that we lack an adequate empirical basis for Monte-Carlo propagation of input distributions, but there are many other options available (e.g., Info-Gap analysis, uncertainty set analysis, consideration of alternative assumption sets and their implied constraints on possible risks, etc.) that could at least provide useful bounds on the plausible risks and on the VOI of reducing uncertainties further (especially, perhaps, on whether the dose-response relation has a threshold – a topic still not settled, despite the pages of discussion.)

6.2a (Volitional uncertainty)

I recommend focusing on traditional value-of-information calculations and decision-analytic approaches to uncertainty analysis. I disagree that a quantitative uncertainty analysis cannot or should not be undertaken, and indeed consider that undertaking one is crucial for providing readers and risk managers with the information needed to inform wise decisions about TCDD. If the mood of the public and of Congress is to be impatient for a decision, then I believe that it would be responsible (though not necessarily popular) to respond with a well-developed decision analysis and quantitative uncertainty analysis (including VOI calculation), with insistence that such quantitative analysis should provide the basis for risk management recommendations. For those who insist that action now is or should be a no-brainer, I would

caution that the data simply do not justify such an attitude: there is a lot of doubt and uncertainty about the probable health consequences (if any) of changing current TCDD exposures, and a well-developed quantitative uncertainty analysis is needed to show whether further action is justified in terms of probable health benefits. Perceptions that the science is settled, and that delays amount to mere stalling (voiced in some public comments recently) suggest the extreme importance of better analyzing and communicating the uncertainties and doubts that current data warrant.

Cox comments - Updated August 15, 2010

5.1 Is the weight-of-evidence characterization scientifically justified and clearly described?

The weight of evidence analysis does not address whether TCDD is a human carcinogen *at relevant exposure levels*, which is crucial for effective risk management if the dose-response relation contains a threshold (or zero slope of the dose-response curve at the origin, see p. 5-55), or threshold-like nonlinearities (e.g., due to receptor-mediated pharmacodynamics). The decision to use a linear non-threshold model does not appear to be justified by science, insofar as current understanding of TCDD carcinogenesis appears to be very consistent with a nonlinear mechanism (Simon et al. 2009, www.ncbi.nlm.nih.gov/pubmed/19776211). The characterization of TCDD simply as a human carcinogen may be misleading, compared to a more thorough discussion of uncertainties (e.g., that it is possible that TCDD is a weak human carcinogen at high doses, but the evidence is far from conclusive, and it is at least as likely that TCDD at current exposure levels is not a human carcinogen). The public comments suggest that this basic uncertainty is not being effectively communicated to the public; by this criterion, the weight of evidence characterization is not clearly described. A much clearer characterization of uncertainties is probably essential to inform readers about what the evidence does and does not establish about the carcinogenicity of TCDD.

The discussion of epidemiological evidence for a causal relation between TCDD exposure and increased risk of cancer draws conclusions that seem to be much stronger than the underlying studies justify. For example, the most recent published update from Seveso found that “All cancer incidence did not differ from expectations in any of the contaminated zones” (Pesatori et al., 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2754980/?tool=pmcentrez&report=abstract.) The Hill-type criteria for evidence of “causality” (strength of association, temporality, consistency, gradient, etc., p. 5-6) do not discriminate between causality and false-positive associations due to multiple testing, multiple comparisons, ignored model uncertainty, model specification error, omitted confounders, residual confounding, unmodeled errors-in-explanatory variables, selection of cut points, variable selection and coding, etc. The conclusion that “In summary, EPA finds the available epidemiological information provides strong evidence of an association between TCDD exposure and human cancer that cannot be reasonably attributed to chance or confounding and other types of bias” does not seem well justified by careful analysis and refutation of alternative explanations and sources of spurious associations, such as those just listed. Insofar as most epidemiological studies have been found to be very prone to type 1 errors (e.g., Ottenbacher, 1998, <http://www.ncbi.nlm.nih.gov/pubmed/9554599>) and related sources of bias (e.g., <http://www.ncbi.nlm.nih.gov/pubmed/17186399>), a more careful and critical review of possible non-causal interpretations of the available epidemiological studies is needed.

5.2 Mode of action.

The available data support EPA's conclusions (p. 5-21) that Ahr activation is probably a necessary precursor step for carcinogenic action, and that subsequent steps in the causal pathways leading to cancer (if TCDD causes cancer) are uncertain. However, a default assumption of a low-dose linear, nonthreshold dose-response relation is not justified by these conclusions. To the contrary, an appropriate dose-response relation for receptor-mediated carcinogenic effects at low doses (with ligand arrivals at receptors being rare events) will typically be sub-linear at low doses (reflecting the sub-linear kinetics for the probability that enough ligands are bound simultaneously to trigger a harmful response). This is not made clear in the discussion of p. 5-53, which only considers equilibrium kinetics, rather than rare-event (non-equilibrium) kinetics appropriate for low levels of exposure. The use of a low-dose nonthreshold modeling approach seems to me to be very questionable, and not consistent with the partial information that is available about the role of Ahr activation. (The assumption that responses are proportional to dose is not well justified by science and does not describe how receptor-mediated responses typically work at low doses.) Thus, while EPA has focused relatively hard on PBPK modeling, I think that better modeling of nonlinear pharmacodynamics (e.g., Simon et al., 2009, www.ncbi.nlm.nih.gov/pubmed/19776211) is important in order to use beyond policy defaults and exploit the partial – but extremely useful – knowledge of Ahr-mediated responses now available. Using a low-dose linear non-threshold default for TCDD does not seem consistent with applying relevant biological knowledge to create more realistic risk models and risk estimates.

The discussion of low-dose linearity (cf p. 5-65) does not seem to me to be clear, well-focused, and technically sound. For example, if Ahr is already being activated by non-TCDD exposures, then the incremental risk from TCDD could presumably be zero (in contrast to the Crump argument that incremental contributions are approximately linear). On the other hand, if Ahr is not already being activated by non-TCDD exposures, then the use of rare-event ligand binding kinetics (leading to a sublinear dose-response relation) would seem to be appropriate. The Portier et al., Crump et al., population variability, etc. arguments appealed to in this section do not address the basic need to consider rare-event kinetics. (By rare-event kinetics, I mean the case in which the arrival of ligands at a receptor is well described by a Poisson process, with mean arrivals per unit time proportional to effective dose rate, and the expected number of bound ligands at any time is small compared to the number needed to trigger a response. The probability that a response will be triggered during any time interval is then a sub-linear function of the effective dose rate.)

Overall, it seems to me that current biological knowledge of MoA, while certainly very incomplete (e.g., we do not even know whether TCDD definitely poses a carcinogenic hazard for humans at relevant levels of exposure) justifies a default assumption of zero slope at the origin, with zero incremental risk in the cases where Ahr activation is either already present for other reasons, or is not already present and is not triggered by TCDD exposure. Certainly, such a model deserves at least as much weight in a scientifically-based risk assessment as a default linear no-threshold model.

6.2 (Infeasibility of a comprehensive quantitative uncertainty analysis)

Although a completely comprehensive quantitative uncertainty analysis might indeed be too much to expect, I think that it is both possible and practical to provide readers with much more useful information about uncertainty. A policy maker might reasonably expect the report to provide insight into major uncertainties and questions such as the following:

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- How likely is it that TCDD is not a human carcinogen at current exposure levels? Full discussion of this uncertainty may help to overcome probability neglect and action bias (Patt and Zeckhauser, 2000, <http://www.springerlink.com/content/k47064873365w720/>).
- What is the probability that reducing TCDD exposures would not reduce cancer risk at all, based on recent epidemiological studies and updates such as Pesatori et al., 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2754980/?tool=pmcentrez&report=abstract?
- What is the probability that reducing TCDD exposures would reduce cancer risk by less than 1 excess cancer case per decade (or per year or per century) in the whole US population, under current conditions?
- What is the probability that reducing TCDD exposures would increase cancer risk (e.g., if the dose-response relation is J-shaped or U-shaped)?
- What is the decision-analytic value of information (VoI) from collecting more information on Ahr kinetics and dose-response before making risk management decisions? Although many members of the public believe that it is imprudent and/or morally wrong to delay tighter regulation of TCDD exposures (perhaps reflecting beliefs that TCDD is a potent carcinogen, developmental toxin, etc.) EPA should provide a thorough quantitative decision analysis that makes explicit the current uncertainties and trade-offs and that shows the conditions under which acting now or postponing action are the optimal actions. Without such quantitative analysis, risk management decisions for TCDD will not be adequately informed, and principles other than those of rational decision-making (e.g., the biases discussed in Sunstein and Zeckhauser, 2010, <http://www.hks.harvard.edu/fs/rzeckhau/Sunstein4-6-09.pdf>) may dominate risk management decisions for TCDD. EPA's uncertainty analysis should provide the (decision and management science) scientific basis for improved decision-making. The current decision to, in effect, punt on quantitative uncertainty analysis is not adequate for informing responsible risk management decision and policy-making, and is not justified.

While I agree with EPA that a quantitative uncertainty analysis is challenging, I do not think that it is impractical to undertake one. It may well be true that we lack an adequate empirical basis for Monte-Carlo propagation of input distributions, but there are many other options available (e.g., Info-Gap analysis, uncertainty set analysis, consideration of alternative assumption sets and their implied constraints on possible risks, etc.) that could at least provide useful bounds on the plausible risks and on the VOI of reducing uncertainties further (especially, perhaps, on whether the dose-response relation has a threshold – a topic still not settled, despite the pages of discussion.)

6.2a (Volitional uncertainty)

I recommend focusing on traditional value-of-information calculations and decision-analytic approaches to uncertainty analysis, and also considering the use of multiple models and ensembles to characterize the “deep uncertainty” (Bankes, 2002, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC128595/pdf/pq10u2007263.pdf>; Kleindorfer

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

2008, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1310239###) about the correct dose-response model and resulting risk predictions at realistic exposure levels.

I disagree that a quantitative uncertainty analysis cannot or should not be undertaken; indeed, performing one is crucial for providing readers and risk managers with the information needed to inform wise decision-making about TCDD. If the mood of the public and of Congress is to be impatient, then I believe that it would be responsible (though not necessarily popular) to respond with a well-developed decision analysis and quantitative uncertainty analysis (including VOI calculation), with insistence that such quantitative analysis should provide the basis for risk management recommendations. For those who insist that action now is or should be a no-brainer, I would caution that the data simply do not justify such an attitude: there is a lot of doubt and uncertainty about the probable health consequences (if any) of changing current TCDD exposures, and a well-developed quantitative uncertainty analysis is needed to show whether further action is justified in terms of probable health benefits. Perceptions that the science is settled, and that delays amount to mere stalling (voiced in some public comments recently) suggest the extreme importance of better analyzing and communicating the uncertainties and doubts that current data warrant.

Dr. Elaine Faustman

Preliminary Comments – July 13, 2010

Response to Charge question Section 2—Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis

This reviewer looked for two things in evaluating this review charge. First, did EPA identify a clear and transparent process for their selection of key data sets for DR analysis and second, did they carry through with such an analysis in the Dioxin reanalysis document.

This reviewer finds that EPA did both, defining a clear and transparent process and then conducting their review in their document in a manner consistent with what they said they would do. EPA deserves many kudos with this impressive part of this document.

The EPA states that in order to respond to the NAS report, they will develop a clear and transparent process by 1. Conducting a comprehensive literature review, 2. Publish this literature on the web for comment, 3. Develop a set of criteria for study inclusion and hold a workshop to review both the draft study inclusion criteria and solicit input from the public on both the literature review and study criteria, 4. Develop the final inclusion criteria for both animal and epidemiological studies taking into account the input they received from the above processes, 5. Prepare their final literature collection (Oct 2009), 6. Screen their studies for inclusion using their criteria, and 7. Develop the final list of key cancer and noncancer studies for quantitative dose-response analysis of TCDD.

This reviewer felt that the document represented a very clear identification of their process and identification of studies. Flow diagrams (ES-1 and ES-2) very clearly show how studies were chosen for inclusion. In general this reviewer agrees with these inclusion criteria as appropriate for both the epidemiology and animals studies. This reviewer was amazed by the number of interesting dioxin research papers. This review will become very, very valuable in this very crowded field of research.

This reviewer did note that EPA states that over 2000 studies were available for review for the dioxin assessment. Frequently, in other types of comprehensive reviews the number of studies considered at each point in a flow diagram such as is seen in Figure 2-2 and 2-3 that did not make the compiled list is given as is the methodology for searching for these articles (see articles on Meta analyses). Since this is so large, it might be useful to know how the searches were conducted (separate from the very extensive and laudable public input processes) and also number of articles that were on dioxin but which were not used by the EPA.

The summary tables (such as Table 2-7 for animal bioassay studies and Table 2-7 for studies for noncancer dose response) were very useful in Chapter 2 to provide a detailed but very readable format for the study data. It was noted that for the epidemiological studies that a very clear rationale for selection versus considered but not selected and this was included in the form of Table 2-3. There were not similar tabular summaries for the animal literature however this could have been useful. Even an extra column in Table 2-7 listing by number reference the criteria met or not met by each study.

Appendix B which included a point by point evaluation of what epidemiological studies would be included and excluded was very useful and provided an extremely detailed rationale for why EPA

included the epidemiology studies that we see in this document. Well done comprehensive look at the process and rationale for use in quantitative evaluations.

Appendix G discusses some of the animal studies that were not selected for evaluation of dose response for noncancer endpoints. It describes the rationale by endpoint for a few studies that were not included in the reference dose derivation based on toxicological relevance.

To this reviewer it was helpful however it was also surprising that the rationale for not considering the animal studies listed in Appendix G was the definition of adversity from IRIS rather than including more specific reference to endpoint guideline documents that EPA has for reproductive and developmental, neurotoxicity, etc. The document and rationale would be much stronger if the specific rationale rather than the general definition of adversity is used exclusively. Please cite these very well constructed and considered documents as support for using or not using these endpoints. This becomes especially important when endpoints such as occur during development are considered where conditions of reversibility are weighed differently than when endpoints are considered in general toxicity.

This reviewer did note that all these activities in response to the NAS comments were designed to more clearly identify studies and not necessarily to integrate these studies. Also this reviewer noted that in the papers for both the epi as well as the animal studies, the authors of these papers but not EPA frequently cite references to a broader category of compounds than just TCDD. For example in numerous papers reference to PCB like compounds is made however with the criteria that EPA has designed for identification of key data sets for dose-response analysis, that there is definitely the very strong focus on ability to use the data/study for quantitative comparisons only on TCDD yet the importance of qualitative similarities, etc. would also significantly inform the document. Later integrative comparisons are made in section 5 for cancer endpoints and the discussions of weight of evidence, however similar considerations of the qualitative body of research findings for noncancer endpoints was not clearly delineated in the document. EPA would not lose this strong and related body of literature that is missing in the point by point evaluation approach for noncancer endpoints if a similar weight of evidence qualitative consideration was included. This becomes important again when the quantitative comparisons are made across species in figures such as ES-5. This reviewer had to “keep in mind” the larger body of literature that would add additional weight to these dose considerations.

Dr. Scott Ferson

Preliminary Comments – July 9, 2010

Responses to charge questions concerning section 6

The arguments in section 6 are coherent and fairly reasonable, although they overstate some issues and underserve some others. This section carefully considers the surprisingly detailed criticisms from the National Academy of Science (NAS) review committee of the 2003 Reassessment concerning the need for quantitative uncertainty analysis. EPA has declined many if not most of the particular suggestions of NAS about uncertainty, and it argues that undertaking the suggested analyses would necessitate further fundamental research in uncertainty quantification. Although I find some of its arguments to be compelling, I also wonder whether EPA has really been responsive to the central criticism about uncertainty. Despite my own strong disposition in favor of quantitative uncertainty analysis in general, it is possible to conclude the agency's judgments on this matter have been thoughtful and defensible.

The following are several comments aimed at improving the text.

The meaning of the phrase 'epistemic uncertainty' given on page 6-5 is plainly incorrect. Epistemic uncertainty is the uncertainty that arises from imperfect knowledge such as from limitations on the amount or quality of data available or deficiencies in our causal understanding about a system. It is not true that a quantity about which there is epistemic uncertainty is necessarily fixed. Although I can see how one might come to this mistaken impression, as far as I know, no researchers use the phrase to imply that the underlying quantity has no variability (although all would admit that this could be the case given our ignorance about it). This mistake echoes in a couple of other places throughout this section.

There is some strange text on the subject of dependence. Lines 30-32 on page 6-5 and section 6.1.3.3 are also incorrect that the "[i]ssues involving...epistemic and aleatory uncertainty translate into issues of dependence". This is just wrong (even under their unusual definition of 'epistemic'). Likewise, the last paragraph on page 6-7 extending onto the next page should be rewritten. The example is reasonable and important, but the discussion about it is confused. The first sentence is incorrect. The uncertainty mentioned in the second sentence may be epistemic, but the sentence is erroneous in its claim. In the following sentences, the words 'variable' and 'fixed' (or 'constant') should be used rather than 'aleatoric' and 'epistemic'. I believe it is nonsense to say that a kinetic constant is "completely correlated across individuals". It's not correlated; it is invariant. This case is not an example of a dependence issue. There is no correlation between a distribution and a fixed quantity (even if it's uncertain). Correlation is defined between *varying* quantities. If the number is fixed, whether or not we know what it is, then you cannot say it's correlated with anything. The authors may have come to this twisted language because they're thinking of the uncertainties in terms of how they might plan to quantitatively characterize them in a Monte Carlo simulation (repeatedly selecting a random deviate for the kinetic constant but assigning it to every individual). Of course, variables such as body fat, age, and smoking, on the other hand, can and do exhibit correlations that definitely should be accounted for in the quantitative assessments. Likewise, the constancy of particular quantities about which we may not know the precise value is also important to keep track of. These two issues should be untangled and discussed in a less confusing way.

It is not clear to me what the authors take to be the difference between epistemic uncertainty and what they call 'cognitive uncertainty'. It seems that the latter phrase was introduced because the meaning of 'epistemic uncertainty' had been misunderstood. Normally, the phrase 'cognitive

uncertainty' would refer to an individual person's uncertainty about the validity of the results of his or her own information processing. The assertion that cognitive uncertainty may be represented by probability (i.e., by precise probability measures) is unnecessary and may be misleading. In fact, researchers in human cognition and neuroscience have shown that humans process this kind of uncertainty (which they often call 'ambiguity') separately and differently from what we think of as probability or frequentist risk (Hsu et al. 2005; Glimcher 2003). I suggest that the section can omit the phrase 'cognitive uncertainty' altogether and use in its place 'epistemic uncertainty'. There are slight differences between the two ideas (e.g., epistemic uncertainty could be shared by members in a group, whereas cognitive uncertainty is always personal), but these appear to be unimportant in this context.

The assertion (on line 10 of page 6-5) that the frequentist and Bayesian interpretations are not mutually exclusive may be misleading. They are mutually exclusive in the sense that it would be improper to mix and match components of each into an analysis. I believe it would be appropriate to omit the clause with the phrase 'mutually exclusive', although it is surely fair to say that subjective probabilities can and do track relative frequencies.

Section 6.1.3.2 starting on page 6-6 discusses a way to address uncertainty for sample data. This Spartan treatment does not mention that sampling uncertainty is not the only kind of uncertainty that can be associated with data, nor that it may not even be the largest kind of uncertainty. Mensurational uncertainty (including the plus-minus part of a measurement, and censoring) may be more important. In some cases, the family or shape of the marginal distribution may be unknown, which is a kind of model uncertainty. As suggested on page 6-35, such uncertainties can be significant. The section suggests only a resampling approach to expressing the uncertainty, but fails to mention the often severe limitations of such approaches, and says nothing about what one might do if there is no relevant sample data.

The first paragraph of section 6.1.3.4 seems to be saying that one can sometimes express model uncertainty as parametric uncertainty, which simplifies its handling. This could be said rather more plainly. It would be helpful to mention that this trick cannot always be used (as when the possible models cannot be listed). It might also be especially helpful to mention that this trick is not so much a way to propagate model uncertainty as a way to sweep it under the rug. Model averaging, including Bayesian model averaging, erases model uncertainty in the same way that averaging variable quantities erases their variation. Bayesian model averaging is mentioned several times in the document, including on page 6-36, lines 3ff. I believe that this method has substantial disadvantages that may disqualify it for consideration here, even as an "exotic" method. Having said this, I would hasten to emphasize that addressing model uncertainty is often useful, and could be useful here as well despite the pessimism of 6.4.2.8. Even a restricted sensitivity analysis, although clearly not comprehensive, can still be informative.

Section 6.1.3.6 starting on page 6-9 might also mention *graphs*, and other traditional communication tools other than correlation indices.

Overall, I think the arguments in section 6 are fairly reasonable, or at least tenable. Although I cannot completely subscribe to the document's conclusion that a reasonably comprehensive quantitative uncertainty analysis is not yet *possible* owing to a lack of models on which to hang the analysis and unavailability of key empirical evidence, I agree that a serious effort in this direction requires further development that may not be justified on *practicality* grounds in this case.

EPA may be overstating the argument a bit, and some text should perhaps be softened. The assertion "Data are the ultimate arbiter of whether quantitative uncertainty analysis with uncertainty

factors, as currently envisioned, has sufficient evidentiary support” (page 6-21, lines 12-14) flies in the face of how uncertainty analyses are normally conceived. Of course, the absence of data is never a substantive reason *not* to conduct an uncertainty analysis; it is the reason *to* do one.

Nevertheless, I agree that an uncertainty analysis is not an absolute good. If the answer is already clear, it can be a waste of time and other resources. If it is used strategically to avoid rendering a proper decision, it can be counterproductive. If it is done poorly, or without appeal to available evidence from the real world, it can be misleading. Surely, if it is worth doing, it is worth doing well and doing something well can be resource-intensive. The idea, mentioned in footnote 66 on page 6-20, of arbitrarily converting uncertainty factors to independent lognormal random variables in a scattered attempt to mount a quantitative uncertainty analysis would entail a suite of unjustified and probably untenable assumptions rendering the exercise nearly pointless.

The pessimistic conclusion on page 6-31, line 24, may be a bit strong. Any *estimate* made from data is amenable to a quantitative uncertainty analysis so, if you’re measuring anything, you can propagate uncertainties such as mensurational uncertainty, sampling uncertainty, and perhaps even surrogacy uncertainty. I don’t think it’s quite as hard to get quantitative models as the text here seems to suggest. Likewise, the similarly dour conclusion on lines 13-14 of page 6-32 leaves me confused. You could do a sensitivity analysis in this case, couldn’t you? If so, it seems that some kind of uncertainty analysis is clearly possible. The caveat on line 29 of page 6-37 is also overwrought. I think exploring relevant alternative values in a sensitivity analysis could constitute a quantitative uncertainty analysis, even if the exploration is limited.

It is important to keep in mind that, in general, we are not necessarily limited to identifying precise probability distributions for everything that is to be characterized as uncertain (as seems to be suggested on line 30 of page 6-37). Simple intervals about uncertain quantities can support a straightforward, albeit crude, interval analysis that propagates uncertainty about parameters and other model choices to statements about the range of possible results. Similarly, an approach based on interval probabilities, probability boxes, or general imprecise probabilities (Walley 1991) can combine such intervals with precise distributions if they are known for some other inputs, and with structures that are intermediate between coarse intervals and delicate probability distributions when some but incomplete knowledge is available. If the inputs are profoundly uncertain, the results from such analyses are likely to be wide in reflection of these uncertainties. In pretty much all cases, it is possible to be entirely rigorous without necessarily being precise and without completely specifying each probability distribution.

There does not need to be a specified “underlying distribution from which to sample” (page 6-37, line 31) in order to conduct a quantitative uncertainty analysis. I think it is a bit too facile to shrug off a call to characterize and account for important uncertainties in the assessment process on these grounds alone. Even when the uncertainty is volitional, there can be relevant ranges that are interesting to decision makers and stakeholders. In such cases, the analysis may be formally closer to a sensitivity analysis, but some appropriate response is usually possible, if not always practicable. To their credit, EPA has acknowledged the legitimacy of the call and undertaken some efforts in this direction, notably Tables 5-18 and 5-19 (although some kind of graphical summary of the results might have been nicer).

The assertions in section 6.5.2 are rather surprising and questionable. EPA says that uncertainty quantification is an “emerging area in science” and that it is “an area where research could be focused” because “the requisite knowledge does not yet exist” to apply quantitative uncertainty analysis in assessments such as this one for dioxin. The document preemptorily dismisses the utility of “convening a blue-ribbon panel” to identify the proper approach and suggests instead that

“multiple approaches should be encouraged”. Are we to infer that the present review panel shouldn’t try to say what the proper approaches to uncertainty quantification are, even if we think the area is more mature than emerging? Do these statements suggest that the agency will support intramural and extramural research efforts in this direction? And, if not, how can we take these pronouncements seriously? Is it not possible that EPA could benefit from some tech transfer efforts as well as basic research on uncertainty quantification? The paragraph beginning on page 6-42 (line 3) mentions a European idea of bench-test exercises to compare different approaches. It may be worth mentioning that this idea has been implemented in the United States as well (Oberkampf et al. 2004; Ferson et al. 2004).

Minor comments

Page 6-3, line 26: If you want to use the adverb ‘always’, the phrase ‘as a joint distribution’ should be ‘as some characterization of a joint distribution’ to be correct.

Page 6-4, lines 9-12: This text is strange and off-putting. A reader might ask who wrote this and why. It seems opinionated and unnecessary.

Footnote 54: The discussion of alternatives to strict, single-measure probability theory is ham-handed. Neither interval probabilities nor imprecise probabilities (sensu Walley 1991) depart from probability theory; they follow the Kolmogorov axioms. They are motivationally and essentially equivalent to sensitivity analyses, except they do not make use of sampling strategies and can be more comprehensive.

Lines 29-30: It is simply untrue that sensitivity analyses have to be systematic. The word ‘systematic’ might better be ‘comprehensive’ and the word ‘essential’ should be weakened, perhaps to ‘advantageous’.

Page 6-5, lines 4-7 and footnote 55: There seem to be only two axioms mentioned in the text, but Kolmogorov needs three to make probability theory.

Page 6-5: The words ‘aleatoric’ and ‘aleatory’ are both used on this page as (synonymous) adjectives of uncertainty. Actually, in the engineering literature, only ‘aleatory’ is preferred for this use. In any case, please pick one to use.

Page 6-6, line 20: Maybe the last word of the header should be plural.

Line 21: Modern practice has replaced ‘error’ with ‘uncertainty’ in this context.

Footnote 56: You could add ‘or subtracting’ after ‘adding’.

Page 6-7, line 14: I think you should replace ‘The role of dependence modeling’ with ‘Dependence among variables’.

Page 6-8. line 13: Omit the unnecessary fancy after the semicolon.

Lines 15-17: This sentence is nonsense, if I understand what a linear low-dose model is. Parsing the sentence, it seems to say “uncertainty over a...slope...may be quantified, but uncertainty...in slope...cannot be captured” which is self-contradictory. I think what you mean to say is that the linearity assumption is not itself subject to uncertainty quantification.

Page 6-9, line 1: The mathematical symbol x should be italicized, as should all Roman letters throughout the document that represent unknown quantities, i.e., are symbols representing something else rather than names like 'e' the base of the natural logarithms.

Lines 14 and 16: The prefixes 'pseudo' and 'quasi' are not words. Hyphens are needed.\

Page 6-10, lines 29-30: Do you mean '*this* probabilistic language', referring to the word 'likely' in the quoted text?

Page 6-11, line 19: Of course there is no guarantee that linear will be protective.

Page 6-13, line 18: Of course it isn't really apodictic knowledge at all, but rather only an opinion or an assumption. I see your point and agree with it entirely, but perhaps you should use a word other than 'apodictic' here since it's not technically correct.

Page 6-14, lines 33-34: The parenthetical phrase 'volitional uncertainty' should be expanded into a sentence that says what you mean to express. The phrase 'cognitive uncertainty' does not mean anything to me in this context. Perhaps if you expanded it into a sentence too, maybe making it 'epistemic uncertainty' along the way, I would understand what you're trying to say here.

Footnote 62: 'Effective' is misspelled, as is 'cancer'.

Page 6-16, line 5: And it's not really a guarantee of course.

Line 8: The word 'common' should be 'predominant'.

Page 6-17, line 28: The word 'band' should be 'limit'.

Page 6-20, footnote 66: The text starting 'each have an error factor' should be followed by 'of' rather than 'or'.

Page 6-21, line 6: It would be nice to give a hint about what the concerns are.

Page 6-22, line 19: And establishes a concomitant reduction in some UFs?

Line 29: The word 'invokes' should perhaps be 'would require'.

Page 6-23, line 33 and passim: The word 'exotic' is a poor choice that is unnecessarily and transparently loaded.

Page 6-25, line 29: This sentence is ungrammatical.

Page 6-26, line 24 and Figure 6-1: Would it be helpful to draw the 45-degree line on the graph?

Page 6-27, line 10: The word 'epistemic' here is acceptable.

Line 14: The word 'epistemic' here should be replaced by 'fixed across individuals'. And 'is estimated from' should be replaced by 'varies with'. I don't see how half life's estimability from data implies that it is variable.

Page 6-28, lines 1-2: You would need the dependence between the variables to proceed.

Line 9: I think that 'and' should be 'although'.

Page 6-29, line 1-2: There are bounding techniques based on the classical Fréchet inequality that do not require any knowledge of or any assumptions about dependencies.

Line 32: Omit 'to'.

Page 6-33: The example in the text box is great, but the second table seems to say the log-likelihood for LLD is 2.46 and for Hill is 2.16, which would make LLD's larger than Hill's, which contradicts what's said in the text.

Page 6-34, line 4: Shouldn't '*Delivered dose*' be a new bullet?

Line 8: I don't think this statement is true. Perhaps 'statistically more powerful' should be 'typically yield more sensitive'.

Lines 24-25: I don't think it's necessary or helpful to persist with Box's platitude. Model uncertainty is the uncertainty about a model's predictions that arises from doubt about the relevance of that model for making such predictions.

Page 6-37, line 30: This sentence is false. Analytical methods don't sample anything, and analyses based on intervals or imprecise probabilities don't depend on uncertainty distributions (i.e., precise probability distributions).

Page 6-38, line 30 and passim: The adjective 'data driven' needs a hyphen, as it has elsewhere in the document.

Line 23-24: I think this sentence is true, but, again, sampling from a distribution is not the only way to conduct a quantitative uncertainty analysis.

Line 26: What is '(2.a)'?

Page 6-41, line 23: Omitting the word 'extra' would make the sentence more easily understandable.

Line 31: What does 'How Forward?' mean? Is this idiomatic?

The document's reference list is alphabetically arranged, but seems to go from Z back to A again on page R-33.

References

Ferson, S., C.A. Joslyn, J.C. Helton, W.L. Oberkampf and K. Sentz. 2004. Summary from the epistemic uncertainty workshop: consensus amid diversity. *Reliability and Engineering and System Safety* 85: 355-370.

Glimcher, P.W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. The MIT Press.

Hsu, M., M. Bhatt, R. Adolphs, D. Tranel, and C.F. Camerer. 2005. Neural systems responding to

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

degrees of uncertainty in human decision-making. *Science* 310 (5754): 1680–1683.

<http://www.sciencemag.org/cgi/content/abstract/310/5754/1680>

Oberkampf, W.L., J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz and S. Ferson. 2004. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability and Engineering and System Safety* 85: 11-20.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Ferson Comments – Updated August 30, 2010

Ferson’s responses to charge questions concerning section 4

The section seems clear insofar as I am prepared to understand it and it seems to be defensible in that it has followed EPA’s own guidance.

Tony Cox’s doesn’t find the evidence epidemiological compelling that TCDD causes cancer, but his complaint is the Kantian objection that association evidence such as in the Hill criteria doesn’t reveal causality. But this is a complaint about epidemiology and perhaps a good portion of empirical sciences in general.

Harvey Clewell’s criticism about uncertainty in section 3 applies here in section 4 as well.

The description of the uncertainty factors seems to follow EPA guidance and standard practice, but that does mean it necessarily makes sense. I don’t recognize their use as an uncertainty analysis in any meaningful sense. 30 is the UF in the RfD derivation, which is the product of 10 and 3.

It seems that EPA has done the hard work of reviewing the quality and caveats of the evidence that’s necessary for a serious QUA, but has declined to synthesize this work into a proper quantitative analysis. Their guidance says they can do it with UFs.

Ferson’s responses to charge questions concerning section 5

I don’t really have any new comments on this section.

I think I had a premature exclamation reacting to Tony’s criticism of the evidence under the Hill-type criteria. I guess that should have been in this section.

What I said before about section 4 about uncertainty applies here. There’s been an enormous amount of effort, all of which is useful and important, but it just hasn’t been integrated. And I’ll save my thunder for the discussion of section 6.

Ferson’s responses to charge questions concerning section 6

The arguments in section 6 are clearly written, mostly coherent, and perhaps fairly reasonable. I had a lot of preliminary comments, including comments on the document’s wording, some of which is strongly at variance with the literature on uncertainty analysis. So I incorporate those here by reference [see “Elaborated responses...” and “Minor comments” in the following sections].

I was befuddled by the argument EPA used to justify not doing a unified QUA. If the blunt answer to the question of why they didn't is that they couldn't specify precise marginal distributions and dependence functions from existing data, then I reject this reasoning and conclude EPA has not been responsive to the NAS criticism. If you're saying EPA guidance doesn't require a QUA, then I would agree and say that the NAS criticism is perhaps itself unreasonable. Or, if you say that you did do an uncertainty analysis in the form of UFs and the limited sensitivity studies that you've done, then I might agree that's a reasonable position, even if it's old-fashioned or dubious. Or even possibly, if you say that mounting a QUA is a significant and controversial undertaking itself and that doing one shouldn't delay the finalization of the report, that I could get behind just on grounds of practicality in the face of a two-decade-long delay.

Here is the reasoning that I would have to reject: EPA asserts that "Data are the ultimate arbiter of whether quantitative uncertainty analysis ... has sufficient evidentiary support". This flies in the face of how uncertainty analyses are normally conceived. Of course, the absence of data is never a substantive reason *not* to conduct an uncertainty analysis; it is the reason *to* do one.

EPA says it needs an "underlying distribution from which to sample" in order to conduct a quantitative uncertainty analysis. I think this is a misunderstanding. And it is facile to shrug off a call to characterize and account for important uncertainties in the assessment process on these grounds alone. If you can *estimate* the value of a quantity, you should be able to express the uncertainty about the value, otherwise you don't really have a scientific measurement in the first place. And, keep in mind, we are not forced to identify precise probability distributions and dependence functions for everything that is to be characterized as uncertain. Even when the uncertainty is volitional, there can be relevant ranges that are interesting to decision makers and stakeholders. In some cases, the analysis may be formally closer to a sensitivity analysis, but some appropriate response is usually possible, if not always practicable. To their credit, EPA has acknowledged the legitimacy of the call by NAS and undertaken some efforts in this direction,

EPA calls uncertainty analysis an "emerging area in science" and this is inarguably true, but I don't believe it is true that methodological research is necessary for EPA to do anything more comprehensive to respond to NAS's criticism, even if we disallow the use of expert elicitation.

I'm entirely sympathetic to the idea of having analyses be *data-driven*, but it is still possible to do something that's useful, even if it's not precisely distributional. There are a variety of ways to conduct a quantitative uncertainty analysis, even an entirely probabilistic one that obeys the Kolmogorov axioms that require neither a bunch of data nor expert elicitation. I'll provide a list of various ways, with appropriate references [see the bulleted list in the summary of July discussions about section 6]. The list includes simple interval analysis that just propagates the plausible ranges, and the supervaluation approach that uses nested inner and outer intervals, with the inner range representing the values that most everyone considers to be plausible values and the outer range representing conservatively broad ranges. There's also a continuous and unbounded version of nesting intervals in an approach known as info-gap analysis that would be useful if we cannot come up with finite bounds on some of the inputs. You can also propagate bounds on distribution functions, so if you know some but not perfect information about each input variable's distribution or some information about some dependence function between the variables, you can fashion bounds on distribution functions and conveniently propagate them through calculations.

Does using these approaches require EPA to make judgments? Yes, it would, in the same way that developing any analysis requires judgments. This does not mean that analysts would be required to make up stuff or elicit any expert opinion. Does it necessitate a lot of extra work? Not necessarily.

These methods can be simple to develop, and they are mostly computationally trivial. Of course, the more comprehensive it is, the harder it is. But the analysis does not have to be fully comprehensive to be useful.

Nevertheless, I agree that an uncertainty analysis is not an absolute good. If the answer is already clear, it can be a waste of time and resources. I don't support wasting time and resources. If the analysis is done poorly, or without appeal to available evidence from the real world, it can be misleading. If the analysis is used *strategically* to avoid rendering or finalizing a decision that is proper, it can be counterproductive.

The following are synoptic answers to the four charge questions of section 6:

6.1: *Clearly presented and scientifically justified?* Yes, the EPA response is clearly presented, but, no, it is not scientifically justified, although their decision to not do a QUA may be justified on grounds of practicality.

6.2: *Comprehensive QUA is unfeasible?* EPA's claim is untrue. Simple methods, which can be legitimate uncertainty analyses (and perhaps even fully probabilistic) are possible, and useful, and sufficient to respond to NAS' criticism.

6.2a: *Volitional uncertainty?* I think a lot of analysts call it 'decisional uncertainty'. There is a difference between this and other forms of uncertainty, but it is not true that there's nothing that can be done with it.

6.3: *Utility of the sensitivity studies?* The utility of the sensitivity studies is very good, but they are not integrated and they need to be.

So what should they do? Well, we've dumped a lot of issues from our consideration over the last two days of the other sections onto the uncertainty analysis. It might be odd to discharge them all now by suggesting that EPA doesn't need to conduct one. Will a QUA change the outcome of this assessment? Josh Cohen, one of our public commenters who was on the NAS committee, seems to think it would or at least could. I don't know. I'm not sure we can tell without doing one. But maybe EPA's analysts know.

Should the absence of QUA further delay the finalization of this superannuated assessment? I'm not sure that it should. Maybe we should consider this question and weigh our desire for an uncertainty analysis in light of this. We want them to do a better job, but even more we want them to do the job. Are we "past the time for reasonable debate and robust science", as a public commenter said yesterday?

Elaborated responses to charge questions concerning section 6

The arguments in section 6 are coherent and fairly reasonable, although they overstate some issues and underserve some others. This section carefully considers the surprisingly detailed criticisms from the National Academy of Science (NAS) review committee of the 2003 Reassessment concerning the need for quantitative uncertainty analysis. EPA has declined many if not most of the particular suggestions of NAS about uncertainty, and it argues that undertaking the suggested analyses would necessitate further fundamental research in uncertainty quantification. Although I find some of its arguments to be compelling, I also wonder whether EPA has really been responsive to the central criticism about uncertainty. Despite my own strong disposition in favor of

quantitative uncertainty analysis in general, it is possible to conclude the agency's judgments on this matter have been thoughtful and defensible.

The following are several comments aimed at improving the text.

The meaning of the phrase 'epistemic uncertainty' given on page 6-5 is plainly incorrect. Epistemic uncertainty is the uncertainty that arises from imperfect knowledge such as from limitations on the amount or quality of data available or deficiencies in our causal understanding about a system. It is not true that a quantity about which there is epistemic uncertainty is necessarily fixed. Although I can see how one might come to this mistaken impression, as far as I know, no researchers use the phrase to imply that the underlying quantity has no variability (although all would admit that this could be the case given our ignorance about it). This mistake echoes in a couple of other places throughout this section.

There is some strange text on the subject of dependence. Lines 30-32 on page 6-5 and section 6.1.3.3 are also incorrect that the "[i]ssues involving...epistemic and aleatory uncertainty translate into issues of dependence". This is just wrong (even under their unusual definition of 'epistemic'). Likewise, the last paragraph on page 6-7 extending onto the next page should be rewritten. The example is reasonable and important, but the discussion about it is confused. The first sentence is incorrect. The uncertainty mentioned in the second sentence may be epistemic, but the sentence is erroneous in its claim. In the following sentences, the words 'variable' and 'fixed' (or 'constant') should be used rather than 'aleatoric' and 'epistemic'. I believe it is nonsense to say that a kinetic constant is "completely correlated across individuals". It's not correlated; it is invariant. This case is not an example of a dependence issue. There is no correlation between a distribution and a fixed quantity (even if it's uncertain). Correlation is defined between *varying* quantities. If the number is fixed, whether or not we know what it is, then you cannot say it's correlated with anything. The authors may have come to this twisted language because they're thinking of the uncertainties in terms of how they might plan to quantitatively characterize them in a Monte Carlo simulation (repeatedly selecting a random deviate for the kinetic constant but assigning it to every individual). Of course, variables such as body fat, age, and smoking, on the other hand, can and do exhibit correlations that definitely should be accounted for in the quantitative assessments. Likewise, the constancy of particular quantities about which we may not know the precise value is also important to keep track of. These two issues should be untangled and discussed in a less confusing way.

It is not clear to me what the authors take to be the difference between epistemic uncertainty and what they call 'cognitive uncertainty'. It seems that the latter phrase was introduced because the meaning of 'epistemic uncertainty' had been misunderstood. Normally, the phrase 'cognitive uncertainty' would refer to an individual person's uncertainty about the validity of the results of his or her own information processing. The assertion that cognitive uncertainty may be represented by probability (i.e., by precise probability measures) is unnecessary and may be misleading. In fact, researchers in human cognition and neuroscience have shown that humans process this kind of uncertainty (which they often call 'ambiguity') separately and differently from what we think of as probability or frequentist risk (Hsu et al. 2005; Glimcher 2003). I suggest that the section can omit the phrase 'cognitive uncertainty' altogether and use in its place 'epistemic uncertainty'. There are slight differences between the two ideas (e.g., epistemic uncertainty could be shared by members in a group, whereas cognitive uncertainty is always personal), but these appear to be unimportant in this context.

The assertion (on line 10 of page 6-5) that the frequentist and Bayesian interpretations are not mutually exclusive may be misleading. They are mutually exclusive in the sense that it would be improper to mix and match components of each into an analysis. I believe it would be appropriate

to omit the clause with the phrase ‘mutually exclusive’, although it is surely fair to say that subjective probabilities can and do track relative frequencies.

Section 6.1.3.2 starting on page 6-6 discusses a way to address uncertainty for sample data. This Spartan treatment does not mention that sampling uncertainty is not the only kind of uncertainty that can be associated with data, nor that it may not even be the largest kind of uncertainty. Mensurational uncertainty (including the plus-minus part of a measurement, and censoring) may be more important. In some cases, the family or shape of the marginal distribution may be unknown, which is a kind of model uncertainty. As suggested on page 6-35, such uncertainties can be significant. The section suggests only a resampling approach to expressing the uncertainty, but fails to mention the often severe limitations of such approaches, and says nothing about what one might do if there is no relevant sample data.

The first paragraph of section 6.1.3.4 seems to be saying that one can sometimes express model uncertainty as parametric uncertainty, which simplifies its handling. This could be said rather more plainly. It would be helpful to mention that this trick cannot always be used (as when the possible models cannot be listed). It might also be especially helpful to mention that this trick is not so much a way to propagate model uncertainty as a way to sweep it under the rug. Model averaging, including Bayesian model averaging, erases model uncertainty in the same way that averaging variable quantities erases their variation. Bayesian model averaging is mentioned several times in the document, including on page 6-36, lines 3ff. I believe that this method has substantial disadvantages that may disqualify it for consideration here, even as an “exotic” method. Having said this, I would hasten to emphasize that addressing model uncertainty is often useful, and could be useful here as well despite the pessimism of 6.4.2.8. Even a restricted sensitivity analysis, although clearly not comprehensive, can still be informative.

Section 6.1.3.6 starting on page 6-9 might also mention *graphs*, and other traditional communication tools other than correlation indices.

Overall, I think the arguments in section 6 are fairly reasonable, or at least tenable. Although I cannot completely subscribe to the document’s conclusion that a reasonably comprehensive quantitative uncertainty analysis is not yet *possible* owing to a lack of models on which to hang the analysis and unavailability of key empirical evidence, I agree that a serious effort in this direction requires further development that may not be justified on *practicality* grounds in this case.

EPA may be overstating the argument a bit, and some text should perhaps be softened. The assertion “Data are the ultimate arbiter of whether quantitative uncertainty analysis with uncertainty factors, as currently envisioned, has sufficient evidentiary support” (page 6-21, lines 12-14) flies in the face of how uncertainty analyses are normally conceived. Of course, the absence of data is never a substantive reason *not* to conduct an uncertainty analysis; it is the reason *to* do one.

Nevertheless, I agree that an uncertainty analysis is not an absolute good. If the answer is already clear, it can be a waste of time and other resources. If it is used strategically to avoid rendering a proper decision, it can be counterproductive. If it is done poorly, or without appeal to available evidence from the real world, it can be misleading. Surely, if it is worth doing, it is worth doing well and doing something well can be resource-intensive. The idea, mentioned in footnote 66 on page 6-20, of arbitrarily converting uncertainty factors to independent lognormal random variables in a scattered attempt to mount a quantitative uncertainty analysis would entail a suite of unjustified and probably untenable assumptions rendering the exercise nearly pointless.

The pessimistic conclusion on page 6-31, line 24, may be a bit strong. Any *estimate* made from data is amenable to a quantitative uncertainty analysis so, if you're measuring anything, you can propagate uncertainties such as mensurational uncertainty, sampling uncertainty, and perhaps even surrogacy uncertainty. I don't think it's quite as hard to get quantitative models as the text here seems to suggest. Likewise, the similarly dour conclusion on lines 13-14 of page 6-32 leaves me confused. You could do a sensitivity analysis in this case, couldn't you? If so, it seems that some kind of uncertainty analysis is clearly possible. The caveat on line 29 of page 6-37 is also overwrought. I think exploring relevant alternative values in a sensitivity analysis could constitute a quantitative uncertainty analysis, even if the exploration is limited.

It is important to keep in mind that, in general, we are not necessarily limited to identifying precise probability distributions for everything that is to be characterized as uncertain (as seems to be suggested on line 30 of page 6-37). Simple intervals about uncertain quantities can support a straightforward, albeit crude, interval analysis that propagates uncertainty about parameters and other model choices to statements about the range of possible results. Similarly, an approach based on interval probabilities, probability boxes, or general imprecise probabilities (Walley 1991) can combine such intervals with precise distributions if they are known for some other inputs, and with structures that are intermediate between coarse intervals and delicate probability distributions when some but incomplete knowledge is available. If the inputs are profoundly uncertain, the results from such analyses are likely to be wide in reflection of these uncertainties. In pretty much all cases, it is possible to be entirely rigorous without necessarily being precise and without completely specifying each probability distribution.

There does not need to be a specified "underlying distribution from which to sample" (page 6-37, line 31) in order to conduct a quantitative uncertainty analysis. I think it is a bit too facile to shrug off a call to characterize and account for important uncertainties in the assessment process on these grounds alone. Even when the uncertainty is volitional, there can be relevant ranges that are interesting to decision makers and stakeholders. In such cases, the analysis may be formally closer to a sensitivity analysis, but some appropriate response is usually possible, if not always practicable. To their credit, EPA has acknowledged the legitimacy of the call and undertaken some efforts in this direction, notably Tables 5-18 and 5-19 (although some kind of graphical summary of the results might have been nicer).

The assertions in section 6.5.2 are rather surprising and questionable. EPA says that uncertainty quantification is an "emerging area in science" and that it is "an area where research could be focused" because "the requisite knowledge does not yet exist" to apply quantitative uncertainty analysis in assessments such as this one for dioxin. The document peremptorily dismisses the utility of "convening a blue-ribbon panel" to identify the proper approach and suggests instead that "multiple approaches should be encouraged". Are we to infer that the present review panel shouldn't try to say what the proper approaches to uncertainty quantification are, even if we think the area is more mature than emerging? Do these statements suggest that the agency will support intramural and extramural research efforts in this direction? And, if not, how can we take these pronouncements seriously? Is it not possible that EPA could benefit from some tech transfer efforts as well as basic research on uncertainty quantification? The paragraph beginning on page 6-42 (line 3) mentions a European idea of bench-test exercises to compare different approaches. It may be worth mentioning that this idea has been implemented in the United States as well (Oberkampf et al. 2004; Ferson et al. 2004).

Minor comments

Page 6-3, line 26: If you want to use the adverb ‘always’, the phrase ‘as a joint distribution’ should be ‘as some characterization of a joint distribution’ to be correct.

Page 6-4, lines 9-12: This text is strange and off-putting. A reader might ask who wrote this and why. It seems opinionated and unnecessary.

Footnote 54: The discussion of alternatives to strict, single-measure probability theory is ham-handed. Neither interval probabilities nor imprecise probabilities (sensu Walley 1991) depart from probability theory; they follow the Kolmogorov axioms. They are motivationally and essentially equivalent to sensitivity analyses, except they do not make use of sampling strategies and can be more comprehensive.

Lines 29-30: It is simply untrue that sensitivity analyses have to be systematic. The word ‘systematic’ might better be ‘comprehensive’ and the word ‘essential’ should be weakened, perhaps to ‘advantageous’.

Page 6-5, lines 4-7 and footnote 55: There seem to be only two axioms mentioned in the text, but Kolmogorov needs three to make probability theory.

Page 6-5: The words ‘aleatoric’ and ‘aleatory’ are both used on this page as (synonymous) adjectives of uncertainty. Actually, in the engineering literature, only ‘aleatory’ is preferred for this use. In any case, please pick one to use.

Page 6-6, line 20: Maybe the last word of the header should be plural.

Line 21: Modern practice has replaced ‘error’ with ‘uncertainty’ in this context.

Footnote 56: You could add ‘or subtracting’ after ‘adding’.

Page 6-7, line 14: I think you should replace ‘The role of dependence modeling’ with ‘Dependence among variables’.

Page 6-8, line 13: Omit the unnecessary fancy after the semicolon.

Lines 15-17: This sentence is nonsense, if I understand what a linear low-dose model is. Parsing the sentence, it seems to say “uncertainty over a...slope...may be quantified, but uncertainty...in slope...cannot be captured” which is self-contradictory. I think what you mean to say is that the linearity assumption is not itself subject to uncertainty quantification.

Page 6-9, line 1: The mathematical symbol x should be italicized, as should all Roman letters throughout the document that represent unknown quantities, i.e., are symbols representing something else rather than names like ‘e’ the base of the natural logarithms.

Lines 14 and 16: The prefixes ‘pseudo’ and ‘quasi’ are not words. Hyphens are needed.\

Page 6-10, lines 29-30: Do you mean ‘*this* probabilistic language’, referring to the word ‘likely’ in the quoted text?

Page 6-11, line 19: Of course there is no guarantee that linear will be protective.

Page 6-13, line 18: Of course it isn't really apodictic knowledge at all, but rather only an opinion or an assumption. I see your point and agree with it entirely, but perhaps you should use a word other than 'apodictic' here since it's not technically correct.

Page 6-14, lines 33-34: The parenthetical phrase 'volitional uncertainty' should be expanded into a sentence that says what you mean to express. The phrase 'cognitive uncertainty' does not mean anything to me in this context. Perhaps if you expanded it into a sentence too, maybe making it 'epistemic uncertainty' along the way, I would understand what you're trying to say here.

Footnote 62: 'Effective' is misspelled, as is 'cancer'.

Page 6-16, line 5: And it's not really a guarantee of course.

Line 8: The word 'common' should be 'predominant'.

Page 6-17, line 28: The word 'band' should be 'limit'.

Page 6-20, footnote 66: The text starting 'each have an error factor' should be followed by 'of' rather than 'or'.

Page 6-21, line 6: It would be nice to give a hint about what the concerns are.

Page 6-21, lines 12-14: NAS was not suggesting that EPA use the *uncertainty factors* approach to mount an uncertainty analysis, but rather a more modern approach.

Page 6-22, line 19: And establishes a concomitant reduction in some UFs?

Line 29: The word 'invokes' should perhaps be 'would require'.

Page 6-23, line 33 and passim: The word 'exotic' is a poor choice that is unnecessarily and transparently loaded.

Page 6-25, line 29: This sentence is ungrammatical.

Page 6-26, line 24 and Figure 6-1: Would it be helpful to draw the 45-degree line on the graph?

Page 6-27, line 10: The word 'epistemic' here is acceptable.

Line 14: The word 'epistemic' here should be replaced by 'fixed across individuals'. And 'is estimated from' should be replaced by 'varies with'. I don't see how half life's estimability from data implies that it is variable.

Page 6-28, lines 1-2: You would need the dependence between the variables to proceed.

Line 9: I think that 'and' should be 'although'.

Page 6-29, line 1-2: There are bounding techniques based on the classical Fréchet inequality that do not require any knowledge of or any assumptions about dependencies.

Line 32: Omit 'to'.

Page 6-33: The example in the text box is great, but the second table seems to say the log-likelihood for LLD is 2.46 and for Hill is 2.16, which would make LLD's larger than Hill's, which contradicts what's said in the text.

Page 6-34, line 4: Shouldn't '*Delivered dose*' be a new bullet?

Line 8: I don't think this statement is true. Perhaps 'statistically more powerful' should be 'typically yield more sensitive'.

Lines 24-25: I don't think it's necessary or helpful to persist with Box's platitude. Model uncertainty is the uncertainty about a model's predictions that arises from doubt about the relevance of that model for making such predictions.

Page 6-37, line 30: This sentence is false. Analytical methods of propagation (convolution) don't "sample" anything, and analyses based on intervals or imprecise probabilities don't depend on uncertainty "distributions" (i.e., precise probability distributions).

Page 6-38, line 30 and passim: The adjective 'data driven' needs a hyphen, as it has elsewhere in the document.

Line 23-24: I think this sentence is true, but, again, sampling from a distribution is not the only way to conduct a quantitative uncertainty analysis.

Line 26: What is '(2.a)'?

Page 6-41, line 23: Omitting the word 'extra' would make the sentence more easily understandable.

Line 31: What does 'How Forward?' mean? Is this idiomatic?

The document's reference list is alphabetically arranged, but seems to go from Z back to A again on page R-33.

References

- Aughenbaugh, J. M., and C.J.J. Paredis 2007. Probability bounds analysis as a general approach to sensitivity analysis in decision making under uncertainty. *SAE 2007 Transactions Journal of Passenger Cars: Mechanical Systems, (Section 6)* 116: 1325-1339, SAE International, Warrendale, Pennsylvania. http://www.srl.gatech.edu/Members/jaughenbaugh/papers_presentations/2007-01-1480.pdf
- Ben-Haim, Y. 2006. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*. 2nd edition, Academic Press, London.
- Davidovitch, L., R. Stoklosa, J. Majer, A. Nietrzeba, P. Whittle, K. Mengersen, Y. Ben-Haim. 2009. Info-gap theory and robust design of surveillance for invasive species: The case study of Barrow Island. *Journal of Environmental Management* 90(8): 2785-2793.
- Dixon, W.J. 2007. *The use of Probability Bounds Analysis for Characterising and Propagating Uncertainty in Species Sensitivity Distributions*. Arthur Rylah Institute for Environmental Research, Technical Report Series No. 163, Department of Sustainability and Environment. Victoria, Australia [http://www.land.vic.gov.au/CA256F310024B628/0/F11A9B711262E4C3CA2573840014660E/\\$File/ARI+Technical+Report+163+-+The+use+of+probability+bounds+analysis+for+characterising+and+propagating+uncertainty+in+species+sensitivity+distributions.pdf](http://www.land.vic.gov.au/CA256F310024B628/0/F11A9B711262E4C3CA2573840014660E/$File/ARI+Technical+Report+163+-+The+use+of+probability+bounds+analysis+for+characterising+and+propagating+uncertainty+in+species+sensitivity+distributions.pdf)

- EPA (U. S. Environmental Protection Agency) 2007. Region 6 Superfund Program: Calcasieu Estuary Initiative. <http://www.epa.gov/earth1r6/6sf/sfsites/calcnit.htm>, accessed July 2009
- Ferson, S. 2002. *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers*. Lewis Publishers, Boca Raton, Florida.
- Ferson, S. and T.F. Long. 1995. Conservative uncertainty propagation in environmental risk assessments. *Environmental Toxicology and Risk Assessment, Third Volume, ASTM STP 1218*, J.S. Hughes, G.R. Biddinger and E. Mones (eds.), American Society for Testing and Materials, Philadelphia, pp. 97–110.
- Ferson, S., V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. 2003. *Constructing Probability Boxes and Dempster-Shafer Structures*. SAND2002-4015. Sandia National Laboratories, Albuquerque, NM.
- Ferson, S., C.A. Joslyn, J.C. Helton, W.L. Oberkampf and K. Sentz. 2004. Summary from the epistemic uncertainty workshop: consensus amid diversity. *Reliability and Engineering and System Safety* 85: 355-370.
- Glimcher, P.W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. The MIT Press.
- Hall, J.W. and H. Harvey 2009. Decision making under severe uncertainty for flood risk management: a case study of info-gap robustness analysis. *Proceedings of the Eighth International Conference on Hydroinformatics*, January 12-16, 2009, Concepcion, Chile.
http://www.floodsite.net/html/partner_area/project_docs/T20-08-05-hic-infogap.pdf
- Hsu, M., M. Bhatt, R. Adolphs, D. Tranel, and C.F. Camerer. 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310 (5754): 1680–1683.
<http://www.sciencemag.org/cgi/content/abstract/310/5754/1680>
- Karanki, D.R., H.S. Kushwaha, A.K. Verma, and S. Ajit. 2009. Uncertainty analysis based on probability bounds (p-box) approach in probabilistic safety assessment. *Risk Analysis* 29(5):662-75.
- Minnery, J.G., J.G. Jacangelo, L.I. Boden, D.J. Vorhees and W. Heiger-Bernays. 2009. Sensitivity analysis of the pressure-based direct integrity test for membranes used in drinking water treatment. *Environmental Science and Technology* 43(24): 9419–9424.
- Montgomery, V. 2009. *New Statistical Methods in Risk Assessment by Probability Bounds*. Ph.D. dissertation, Durham University, UK. <http://maths.dur.ac.uk/stats/people/fc/thesis-VM.pdf>
- Moore, R.E. 1966. *Interval Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.
- Neumaier, A. 1990. *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge.
- Oberkampf, W.L., J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz and S. Ferson. 2004. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability and Engineering and System Safety* 85: 11-20.
- Regan, H.M., B.K. Hope and S. Ferson. 2002a. Analysis and portrayal of uncertainty in a food-web exposure model. *Human and Ecological Risk Assessment* 8: 1757 - 1777.
- Regan, H.M., B.E. Sample and S. Ferson. 2002b. Comparison of deterministic and probabilistic calculation of ecological soil screening levels. *Environmental Toxicology and Chemistry* 21: 882-890.
- Regan, H.M., Y. Ben-Haim, B. Langford, W.G. Wilson, P. Lundberg, S.J. Andelman, and M.A. Burgman 2005. Robust decision-making under severe uncertainty for conservation management. *Ecological Applications* 15(4):1471-1477.
- Rout, T.M., C.J. Thompson and M.A. McCarthy. 2009. Robust decisions for declaring eradication of invasive species. *Journal of Applied Ecology* 46: 782–786.
- Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.
- Yokomizo, H., H.P. Possingham, M.B. Thomas, and Y.M. Buckley. 2009. Managing the impact of invasive species: the value of knowing the density-impact curve. *Ecological Applications* 19(2): 376-386.

Dr. Jeffrey Fisher

Preliminary Comments – July 8, 2010

Chapter 3.

3.1a—I prefer to use whole blood as the as the dosimeter in this situation instead of a whole body calculation. Future refinements in dosimetry are easier. The blood compartment has been used for many chemicals and drugs as a dosimeter because experimental data are collected from this compartment. Thus one can directly compare model predictions with data derived calculations or measurements.

3.1b—The justification or judgment call for using the Emond et al. model is fairly strong overall. Initially I looked for specific text that stated how the model would be used, which would then dictate why this model was selected. That is, for the intended purposes, this model was more robust than some models and simpler than others and contained sufficient details for important biological determinants deemed important by the Agency.

3.1c—The modifications were minor. Adjusting for both blood volume and urine excretion based on the blood concentration are ok. Not sure if I followed the blood volume requirement. The body blood compartment volume was not accounted for in the original model? For me... seeing the original equation would help.

3.1d—I am unable to answer this question at this time without more time. Deterministic models were used with mean concentration values. Model parameter distributions for humans would be useful.

3.2a—The mouse model is ok, since, apparently, no mouse model was available. I do think a peer review of the model is important and is apparently a normal requirement for models to be used by the Agency. There are many mouse PBPK models in the literature, so this is not a unique endeavor.

3.2b—The performance of the mouse model was moderate to good. Under prediction of urinary excretion may suggest a systemic problem with the model since it was observed with two data sets.

3.2c—Need more time. (rodent to human extrapolation)

3.3 –Need more time (rodent to human extrapolation)

3.4—Sensitivity analysis should be carried out using the dosemetrics of interest (eg., AUC blood) that is used for extrapolation to humans. I found an analysis for only elimination rate.

3.4—Need more time to evaluate.

No comments on Chapter 6 at this time.

Fisher Comments – Updated August 26, 2010

Chapter 3.

3.1a—I prefer to use whole blood as the as the dosimeter in this situation instead of a whole body calculation. Future refinements in dosimetry are easier. The blood compartment has been used for

many chemicals and drugs as a dosimeter because experimental data are collected from this compartment. Thus one can directly compare model predictions with data derived calculations or measurements. Conversion to lipid adjusted value is a calculation in many instances and not determined directly by measuring organics in the fat fraction of blood.

3.1b—The justification or judgment call for using the Emond et al. model is fairly strong overall. Initially I looked for specific text that stated how the model would be used, which would then dictate why this model was selected. This should be the rationale for selection of the model along with the listed justifications. For example, the authors could state that, for the intended purposes, this model was more robust than some models (give examples) and simpler than others (give examples) and contained sufficient details for important biological determinants deemed important by the Agency. The child model for Edmond does not include a pathway for loss of dioxin via the GI tract (secretion into the GI tract and then into feces). The EPA should state why this will not be included as a modification to the Edmond model.

3.1c—The modifications of the Edmond model were minor. Adjusting for both blood volume and urine excretion based on the blood concentration are ok. Showing the original equation and the modification of this equation would be very helpful to understand specifically what was done.

3.1d---A good discussion of the data sets and strengths and weaknesses of the models was present in the report. Biological determinants and statistical information were discussed, relative to the use of the models. The new mouse model and the modified Edmond model should have limited sensitivity completed based on the prior work of Edmond to ensure the model performance is similar to that of Edmond.

3.2a—The mouse model is ok, since, apparently, no mouse model was available. I do think a peer review of the model is important and is apparently a normal requirement for models to be used by the Agency. There are many mouse PBPK models in the literature, so this is not a unique endeavor.

3.2b—The performance of the mouse model was moderate to good. Under prediction of urinary excretion may suggest a systemic problem with the model since it was observed with two data sets. Adjusting the appropriate rate constants would be helpful. This is a minor pathway for excretion and will have minimal influence on predicting the kinetics of dioxin.

3.2c—Methodology for extrapolation to human is fine. Dose extrapolation, accounting for possible pharmacokinetic differences between species, is the best approach.

3.3 –This approach was rationale and acceptable in terms of methods used to estimate intake. Many people in the modeling community perform such extrapolations. In the case of dioxin, with a rather long half life, the input functions can vary less than for chemicals with short half lives. This simply provides a rationale form of interpretation of the biomonitoring data based on what is known about the pharmacokinetics of dioxin.

3.4—Sensitivity analysis should be carried out using the dosimetrics of interest (eg., AUC blood) and reported in the document or cited. Edmond did sensitivity analysis, but his model was modified, so a limited sensitivity analysis based on his work should be carried out. I found an analysis for only the inducible elimination rate (Fig. 3-30). The work may have been done already and not reported.

3.5—Based on my limited knowledge, the daily intake rates for dioxin are ok, with the exception that clusters of people exist who appear to have higher intake rates because of source specific exposures to dioxin.

6.1—The quantitative uncertainty assessment was useful and clearly presented, but I do not think scientifically justified in terms of the conclusions.

6.2 Perhaps, the assessment expectations or the bar was set too high. From an academic perspective, one can conclude that a ‘compressive’ quantitative uncertainty analysis is not feasible. This is probably true for most chemicals and drugs. Short of an omnipotent understanding of dioxin and human health, there are pieces of quantitative uncertainty that can be characterized within the framework of the efforts conducted by the EPA.

6.2a—I can not qualified to comment on this.

6.3—I think these pieces of uncertainty analysis are very important and help clarify focused issues on dioxin.

General Comment: The presentation of dose-response profiles for dioxin using both AhR dependent and AhR independent assumptions was surprising. Abandoning the decades of research on the mechanistic underpinnings of dioxin mediated toxicity (via nuclear receptor binding) is a serious endeavor. The justification for AhR dependent dose response profiles should be substantially enhanced in the document and explained clearly for scientists and the readership at large.

Dr. Helen Håkansson

Preliminary Comments – July 8, 2010

RE: EPA's reanalysis of key issues related to dioxin toxicity and response to NAS comments; SAB Meeting July 13-15, 2010, Washington DC

Preliminary comments to the Charge questions

Considering the length and complexity of the documents to be analysed in relation to the time available between receipt of the information and the meeting event I like to stress that my responses has to be regarded as preliminary and broad. More detailed comments covering all aspects of the evaluation will need additional work and time.

General

- 1.1 In my view, the draft Response to NAS Comments is comprehensive and addresses the three key recommendations in a clear and logical way.
- 1.2 I am aware of a couple of other experimental studies that might meet the criteria of being included in the evaluation. At this stage I cannot say whether these studies will have a significant impact on the hazard characterization or the dose-response assessment.

Section 2

- 2.1 In general, I think this section addresses the transparency concern raised by NAS. Could be that additional studies could be included
- 2.2 In general, I think the study selection criteria are scientifically justified and clearly described
- 2.3 In general, I think the study selection criteria are applied in a scientifically sound manner.

Section 3

Until now no comments

Section 4

In general, I think the approach outlined in this section is a very good starting point for discussions at the meeting. Identification of critical human and animal studies/end-points, as well as critical exposure matters and use of uncertainty factors are presented and discussed in clear and scientifically sound ways. Could be that additional studies could be included (see point 1.2 above)

Section 5

In general, I think it would be beneficial to the process if the mode-of-action discussion that is placed in this section could be revised to become comprehensive enough to cover all types of toxicities i.e. not only cancer mode of action, and then from there take the discussion further to the weight-of-evidence classification of TCDD according to the 2005 Cancer Guidelines to specifically address the NAS comments. I also think there is need of a more comprehensive description of the AhR-KO mouse phenotype(s) as well as the CA-AhR mouse in order to use the argumentation that AhR activation may progress to cancers in rodents and humans.

I am not convinced that the indicated differences in liver and lung cancer mode of actions (figs 5-2, 5-9, 5-10, and corresponding texts) are real; I think the lung figure, which includes disturbances of the retinoid system, is equally applicable to other tissues e.g liver. Furthermore, there is evidence that more nuclear receptors and their associated endogenous ligands, enzymes and/or binding proteins are modulated by AhR activation and/or repression. I think all of these aspects need to be described in the response to the NAS cancer mode-of-action comments to fully address their concerns about the carcinogenic profile of TCDD. Whether this more comprehensive and general description of the TCDD mode-of-action state-of-the-art should be in section 5 can be discussed.

I am not sure that this section sufficiently clearly addresses the NAS report recommendation that EPA should provide risk estimates and discuss pros and cons involved in the use of both non-linear and linear methods to extrapolate below PODs. While the linear part is comprehensive and convincing in its writing, this is not the case for the non-linear part.

Section 6

Until now no comments

Other comments

It would be useful for the process if it was possible during the meeting to provide some information on the separate EPA-activity, which addresses the TEF/TEQ comments by NAS. It is important that the contact and communication between the TCDD risk characterization and TEF/TEQ processes are open.

Håkansson Comments – Updated September 1, 2010

RE: EPA's reanalysis of key issues related to dioxin toxicity and response to NAS comments; SAB Meeting July 13-15, 2010, Washington DC

Revised Preliminary comments to the Charge questions

Considering the length and complexity of the documents to be analysed in relation to the time available between receipt of the information and the meeting event I like to stress that my responses has to be regarded as preliminary and broad. More detailed comments covering all aspects of the evaluation will need additional work and time.

General

- 1.1 In my view, the draft Response to NAS Comments is comprehensive and addresses the three key recommendations in a clear and logical way; the only exception being the response to the recommended evaluation of a possible non-linear cancer mode-of-action (see below comment under section 5).
- 1.2 I am aware of a couple of other experimental studies that might meet the criteria of being included in the evaluation. At this stage I cannot say whether these studies will have a significant impact on the hazard characterization or the dose-response assessment.

Section 2

- 2.1 In general, I think this section addresses the transparency concern raised by NAS. Could be that additional studies could be included
- 2.2 In general, I think the study selection criteria are scientifically justified and clearly described
- 2.3 In general, I think the study selection criteria are applied in a scientifically sound manner.

Section 3

Until now no comments

Section 4

In general, I think the approach outlined in this section is a very good starting point for discussions at the meeting. Identification of critical human and animal studies/end-points, as well as critical exposure matters and use of uncertainty factors are presented and discussed in clear and scientifically sound ways. Could be that additional studies could be included (see point 1.2 above)

Section 5

In general, I think it would be beneficial to the process if the mode-of-action discussion that is placed in this section could be revised to become comprehensive enough to cover all types of toxicities i.e. not only cancer mode of action, and then from there take the discussion further to the weight-of-evidence classification of TCDD according to the 2005 Cancer Guidelines to specifically address the NAS comments. I also think there is need of a more comprehensive description of the AhR-KO mouse phenotype(s) as well as the CA-AhR mouse in order to use the argumentation that AhR activation may progress to cancers in rodents and humans.

I am not convinced that the indicated differences in liver and lung cancer mode of actions (figs 5-2, 5-9, 5-10, and corresponding texts) are real; I think the lung figure, which includes disturbances of the retinoid system, is equally applicable to other tissues e.g liver. Furthermore, there is evidence

that more nuclear receptors and their associated endogenous ligands, enzymes and/or binding proteins are modulated by AhR activation and/or repression. I think all of these aspects need to be described in the response to the NAS cancer mode-of-action comments to fully address their concerns about the carcinogenic profile of TCDD. Whether this more comprehensive and general description of the TCDD mode-of-action state-of-the-art should be in section 5 can be discussed.

I am not sure that this section sufficiently clearly addresses the NAS report recommendation that EPA should provide risk estimates and discuss pros and cons involved in the use of both non-linear and linear methods to extrapolate below PODs. While the linear part is comprehensive and convincing in its writing, this is not the case for the non-linear part.

Section 6

Until now no comments

Other comments

It would be useful for the process if it was possible during the meeting to provide some information on the separate EPA-activity, which addresses the TEF/TEQ comments by NAS. It is important that the contact and communication between the TCDD risk characterization and TEF/TEQ processes are well coordinated.

I have great difficulties to envision how the scientific knowledge generated during the years from 2003 until now (almost a decade of active science) can be easily (i.e. within a short time-period) integrated in the 2003 USEPA risk assessment draft in a way so that it becomes crystal clear why the two selected human studies are the most appropriate for setting the guidance value. The new knowledge in the areas of developmental toxicology (i.e. mainly the non-cancer end-points) and mode-of-action understanding that has developed during this last decade is critical supporting scientific information lending credibility to the selection of the human non-cancer endpoint-studies. Therefore, I wonder if the process would be simplified if the EPA's reanalysis of key issues related to dioxin toxicity and the response to NAS comments are handled as two separate entities; or maybe this is the plan. If I recall right, NAS do not specifically ask for a new/revised risk assessment. I assume that a strict response to the NAS comments could be accommodated in a report < 30 pages, while an up-to-date revised risk assessment may require more pages.

I realize that these additional comments are beyond the scope of the SAB charge questions and the panel duties however, since the process to me seems to be very open, I anyway take the opportunity to mention these thoughts.

Dr. Russ Hauser

Preliminary Comments – July 7, 2010

Section 2.

2.1: Generally the section was responsive to NAS concerns about transparency and clarity in data set selection for dose-response analysis. Five considerations were used to evaluate the epidemiologic studies. Three inclusion criteria were then used to select studies to use for TCDD quantitative dose-response assessment.

2.2: The criteria and considerations could be further refined and clarified (see specific comments below). Unclear why criteria 1 (study is peer-reviewed) is not one of the first considerations.

2.3: The criteria and considerations were applied in a consistent scientific manner.

Page 2-6:

Consideration #2: Worded awkwardly, mixes terms, etc

- 1) Define 'susceptible' to important biases (very qualitative term)
- 2) Was the text 'control for potential confounding exposures' intentionally worded to only relate to exposures (such as DLC) or can/should it relate more broadly to confounding from exposures and other factors?
- 3) I assume bias arising from study design refers to selection bias, or is it used more broadly for how exposure and outcome are measured and covariate data collected? Rewording may help
- 4) Define what is meant by bias arising from statistical analysis? Unsure bias is correct term. Is this meant to refer to model misspecification?

Consideration #3:

- 1) This is a more general content in regards to risk assessment.

Does the way this is worded preclude inclusion of null studies? Although I'm not an expert in risk assessment methods, when epidemiologists review the literature we consider null studies as providing important information towards the weight of the evidence. This is especially relevant to the thyroid data (Baccarelli) for the non-cancer risk assessment. There are other studies which report null associations. How are/should these be considered in the risk assessment?

Consideration #5:

- 1) Check definition of type II error rate (failing to reject null when alternative is true)

Page 2-7:

Inclusion criteria 2: Define 'primarily' to TCDD. Does this refer quantitatively to TCDD exposure being much higher than DLCs? Or to unique situations where there is only TCDD exposure apart from low background exposure to other DLCs? Or both?

Even in Seveso with high acute exposure to TCDD there was background DLC exposure (exposure to DLCs in 1976 were high compared to today's levels). Thus Seveso may be categorized as very high TCDD in the setting of 'moderate' DLC exposure by today's background levels. Does the way this inclusion criteria is worded preclude the inclusion of a study in a population with background exposure to DLCs (that includes TCCD)? Would such a study be included if the authors measured serum TCDD concentrations and adjusted for other DLCs concentrations? EPA should consider future implications of the inclusion criteria since it may exclude future studies on background TCDD exposure in risk assessment. Rewording may help here...

Hauser comments

page 42

Inclusion criteria 3: Define 'reported' dose.

Page 2-8:

Line 14-15: Not sure if this is counter to what is done with human studies "the studies using the lowest exposures that show effects will typically drive the RfD or OSF"? Appears the RfD for non-cancer was based on human studies with the highest TCDD exposure. For instance, other studies on TSH and dioxins were not included in the risk assessment (some showed no association). Is this appropriate that animal studies use lowest exposures to show effect but human data relied on studies with highest exposures?

Section 4:

4.1: The rationale for the selection of Mocarelli et al (2008) and Baccarelli et al., (2008) is scientifically justified and clearly described. I am unsure how to consider other studies (especially null studies on dioxins and thyroid function) in derivation of RfD. Justification for using change in semen quality and TSH is described in report. Both are adverse but on a continuum rather than dichotomous (e.g., disease/no disease). Therefore, there is the need to clearly justify at what level of change in health endpoint one considers the dose response analysis; attempted in report, e.g., 5uU/mL for TSH and 20% decrease in sperm count.

4.2a.i: Perhaps EPA could also calculate average exposure two additional ways, 1) mean of pulse exposure alone, and 2) 10 yr critical exposure window only.

4.2.a.ii: Reasonable approach to use a 20% decrease in sperm count since driven by data provided in study (must keep in mind relevance of a 20% decrease in sperm count on the population level versus individual level). They may differ, unlike a dichotomous endpoint (e.g. disease or no disease which have more similar individual and population implications).

4.2.b.i: Use of maternal levels was appropriate

4.2.b.ii: Use of decrease of 5 uU/mL TSH was rationalized in report but I would assume a smaller increase in TSH is relevant on a population level (making it more consistent with the way sperm count decline is used in Mocarelli study). For instance, the dose associated with 1 uU/mL increase in TSH is also relevant for infant health on a population level.

Notes: Page 4-3 (line 26): Define 'clinical' in the term 'clinical adverse effects'. Was the intent to medicalize outcome by using clinical?

Hauser Comments – Updated August 25, 2010

Notes on Dioxin Reassessment

July 2010 Meeting

Section 2.

2.1: This section was responsive to NAS concerns regarding the transparency and clarity in data set selection for dose-response analysis. Five considerations were used to evaluate the epidemiologic studies. Three inclusion criteria were then used to select studies to use for TCDD quantitative dose-response assessment.

Minor comments:

2.2: The criteria and considerations could be further refined and clarified (see specific comments below). Unclear why criteria 1 (study is peer-reviewed) is not one of the first considerations.

2.3: The criteria and considerations were applied in a consistent scientific manner.

Page 2-6:

Consideration #2: Worded awkwardly, mixes terms, etc

- 1) Define 'susceptible' to important biases (very qualitative term)
- 2) Was the text 'control for potential confounding exposures' intentionally worded to only relate to exposures (such as DLC) or can/should it relate more broadly to confounding from exposures and other factors?
- 3) I assume bias arising from study design refers to selection bias, or is it used more broadly for how exposure and outcome are measured and covariate data collected? Rewording may help
- 4) Define what is meant by bias arising from statistical analysis? Unsure bias is correct term. Is this meant to refer to model misspecification?

Consideration #3:

- 1) This is a more general comment in regards to risk assessment.

Does the way this is worded preclude inclusion of null studies into the risk assessment paradigm? How should these null studies be considered in the report and risk assessment? A description of the studies not considered further should be included in the report and justification provided for not including them. This needs to be expanded upon in the report and made more specific and clearer.

Consideration #5:

- 1) Check definition of type II error rate (failing to reject null when alternative is true)

Page 2-7:

Inclusion criteria 2: Define 'primarily' to TCDD. Does this refer quantitatively to TCDD exposure being much higher than DLCs? Or to unique situations where there is only TCDD exposure apart from low background exposure to other DLCs? Or both? This was discussed at the July meeting and the report should include a clear description of how 'primarily' is defined.

For instance, even in Seveso with high acute exposure to TCDD there was background DLC exposure (exposure to DLCs in 1976 were high compared to today's levels). Thus Seveso may be categorized as very high TCDD in the setting of 'moderate' DLC exposure by today's background levels. Does the way this inclusion criteria is worded preclude the inclusion of a study in a population with background exposure to DLCs (that includes TCCD)? Would such a study be included if the authors measured serum TCDD concentrations and adjusted for other DLCs concentrations? EPA should consider future implications of the inclusion criteria since it may exclude future studies on background TCDD exposure in future risk assessments. Rewording may help.

Inclusion criteria 3: Define 'reported' dose.

Section 4:

4.1: The rationale for the selection of Mocarelli et al (2008) and Baccarelli et al., (2008) is scientifically justified and clearly described. Justification for using change in semen quality and TSH is described in report. Both should be considered adverse responses, and it must be made

explicit that these are measured on a continuum rather than dichotomous outcomes (e.g., disease/no disease). Therefore, there is the need to clearly justify the amount of change in these endpoints that should be considered in the dose response analysis. The report did this by using a decrement of 5uU/mL for TSH and a 20% decrease in sperm count.

4.2.a.i: It is a reasonable approach to use a 20% decrease in sperm count since it is based on the data provided in study. Again, it is critical to explicitly note the relevance of a 20% decrease in sperm count on the population level versus on the individual level.

4.2.b.i: Use of maternal levels of serum dioxins (TCDD) was appropriate

4.2.b.ii: The decision to base the dose-response analysis on a decrease of 5 uU/mL TSH was appropriately discussed and justified in report. However, as described above it is important to consider that a smaller increase in TSH is relevant on a population level (making it more consistent with the way sperm count decline is used in Mocarelli study). For instance, the dose associated with 1 uU/mL increase in TSH on a population level is relevant for infant health. That is, a change in the population median TSH can have important public health implications for the number of infants in the tail of the distribution. Therefore, changes in TSH less than 5 uU/mL should be considered **adverse on a population level**.

Notes: Page 4-3 (line 26): Define 'clinical' in the term 'clinical adverse effects'. Was the intent to medicalize outcome by using clinical?

Dr. B. Paige Lawrence

Preliminary Comments – July 8, 2010

Below are my preliminary comments on key issues to be discussed in response to charge questions:

General charge questions

After my first read of the document I find it generally logical and clear. EPA has endeavored to be genuinely responsive to NAS concerns, and has laid out a comprehensible rationale for their approach to the response. The document is complete, with excellent and meticulous consideration of the many complex issues surrounding this chemical.

Specific charge questions

Section 2. Transparency and Clarity of Selection of Key Data Sets For Dose-Response Analysis

This section of the document is highly responsive to NAS concerns and suggestions. The criteria for epidemiological and animal studies are clearly and thoroughly presented, as is the rationale for the parameters used to include or exclude studies. The nuanced differences and complicating issues surrounding this subject are presented in a comprehensive and logical manner. In short, this section is scientifically sound.

Lawrence Comments – Updated August 19, 2010

General Charge Questions

1. Is the draft *Response to Comments* clear and logical? Has EPA objectively and clearly presented the three key NRC recommendations?

This reviewer finds the document generally logical. The EPA has been very responsive to NRC recommendations. However, the meaning of some content is not always clear. One aspect of the document that particularly diminishes overall clarity is that the *Response to NAS Comments* is long and dense with lots of jargon; further, in some places, it is quite repetitive. These features, while a necessity of this type of document, at times detract from clarity or make the EPA's logic difficult to discern. To improve the document further, careful editing by EPA staff to better integrate content throughout the document, reduce redundancies, and consolidate the text is recommended. Two examples of editing that would improve clarity are provided here to illustrate approaches to editing the document:

- Topic sentences are sometimes not easily connected to paragraph content. To provide but one example, take the 2nd paragraph of on page xxvii. I read this paragraph five times, but

still cannot identify the three separate additional EPA activities to address additional NAS comments. I can confidently identify two, but not three.

- Syntax and grammar could be improved throughout the document. For example, the word overview is a noun, but is used in the document as a verb (e.g., p. 2-2 line 23).

2. Are there other critical studies that would make a significant impact on the conclusions of the hazard characterization or dose-response assessment of the chronic non-cancer and cancer health effects of TCDD?

The underlying rationale for study exclusion should be better explained, as excellent studies were excluded, but this reviewer is left not understanding why. To be clear, the explanation for study inclusion is evident (although this reviewer does not agree with the criteria that the purity of TCDD needs to have been stated in the paper). On the other hand, why studies were not included is open to considerable speculation by the reader. A solution to this would be for EPA to revise their explanation of study exclusion criteria, either in the text or associated tables. To be clear, this recommendation is focused on improving clarity of the document. It is not apparent that including other studies would significantly alter the dose-response assessment.

Specific Charge Questions

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis

1. Is this Section responsive to the NAS concern about transparency and clarity in data-set selection for dose-response analysis?

This section of the document is highly responsive to NAS concerns and suggestions. The criteria for epidemiological and animal studies are clearly and thoroughly presented, as is the rationale for the parameters used to include studies. The nuanced differences and complicating issues surrounding this subject are presented in a comprehensive and logical manner. Overall, this section sufficiently clear and transparent.

2. Are the epidemiology and animal bioassay study criteria/considerations scientifically justified and clearly described?

In general the study criteria/considerations are scientifically justified and clearly described. What follows is a minor point of concern. In Section 2.4.1.2.1.5.3 (p 2-110; non-cancer epi studies), the way it is currently written, the rationale for excluding the studies by Baccarelli et al (2002, 2004) on the relationship between TCDD and immunological effects rests on a rather weak foundation. The text in this section states: "Interpreting the inverse association between TCDD exposure and IgG levels in terms of clinical significance is not possible." This is predicated on the idea that if plasma IgG levels do not sink down to those measured in immunocompromised individuals, then there is no clinical significance. This reviewer is not confident that current human and animal immunology data would fully support this. It is possible and likely that there are individuals within a population that may not be diagnosed as immune compromised, but whose immune responses fall outside or on the very edges of the range of normal. However, including these studies would not likely change the outcome of EPA's dose response analyses.

3. Has EPA applied the epidemiology and animal bioassay study criteria/considerations in a scientifically sound manner? If not, please identify and provide a rationale for alternative approaches.

In general, study selection criteria were applied in a scientifically sound manner. As noted above, data set selection could be further justified by editing the text. For example, edits could be made to make it clearer to readers why certain studies were excluded. To be clear, this suggestion does not mean a different approach is needed, but that the approach used should be explained a bit better.

Section 4. Reference Dose

Because the charge questions for Section 4 are rather lengthy, with extensive preambles, they are not restated here. Please refer to the charge questions for context, if it is needed.

4.1: The rationale for the selection of Mocarelli et al (2008) and Baccarelli et al (2008) is and clearly described and the justification is based on sound science. However, as noted in comments on Section 2, EPA needs to do a better job within the document of giving the reader a more transparent and thorough explanation of why other studies were excluded.

4.2.a.i: EPA's approach to identifying the exposure window is clearly explained. There are other ways in which this could be approached; however, it is not certain that using other approaches would impact the outcome.

4.2.a.ii: The designation was clearly justified.

4.2.b.i: Use of maternal levels was appropriate

4.2.b.ii: The decision to use 5 µU TSH /mL was explained well in the Report. A minor concern is that smaller changes in TSH are possibly relevant on a population-wide level, but this concept is generally ignored in the Report.

4.3. (No comment on uncertainty factors at this time)

4.4. This reviewer is one-the-fence regarding the question of whether EPA's choice to not consider biochemical endpoints as potential critical effects for derivation of the RfD for TCDD is scientifically justified. In part, this stems from a belief that these biochemical measures are critically important for understanding how TCDD adversely impacts normal physiological processes. Another factor that contributes to difficulty answering this question is the lack of clarity regarding the exclusion of particular studies (see Section 2).

4.5 EPA's approach for averaging internal blood TCDD concentrations is clear and logical.

4.6 and 4.7 No comment

4.8. The qualitative discussion of the uncertainty in the RfD seems adequately justified. This section could be edited to improve clarity; maybe even use bullet points to highlight and separate key points and/or link with information in other sections (e.g. 6) to make it easier for the reader to appreciate this section of the Report.

Dr. Michael Luster

Preliminary Comments – July 7, 2010

Sec 2

2.1 Yes, overall EPA has been responsive to NAS concerns about transparency and clarity. These responses were taken in the form of request for public comments (e.g., publications of Dx literature), workshops (dose response assessments) and in the responses to the NAS detailed in this document. As I indicate in response to questions in sec.4, I think it would be useful to provide some additional information to help justify the 2 critical studies (Baccarelli et al and Mocarelli et al) that were used to establish the PoD. Also suggest including in Summary section a clarification of why the Emond model is not very reproducible for mice.

2.2 and 2.3: For the most part, YES although I have some minor suggestions for improvement and clarification:

- A. In most cases the rationale for the use of a study, the basis for the PoD selection, or model etc is justified on a clear scientific basis. In several instances, however, the basis for a decision is justified by the statement that it is 'consistent with EPA criteria or policy'. A sentence on the scientific basis of the policy is recommended.
- B. Regarding study inclusion criteria (Fig ES-2): I think additional explanation is required. I assume there were many studies that got to the level of the 4 considerations depicted in the flow chart and most of these already qualified for the 1st and 2nd criteria since these would be part of acceptance for most peer-reviewed publications. Thus, I suspect most of the studies were rejected based upon steps 3 and 4 which are hard to quantify. Better to provide explanation on what is meant by 'consistent toxicological practices' and 'outside normal range of variability' The former I thought would have been identified in the earlier exclusion criteria. Number 4 criteria is particular difficult to understand since there are no accepted normal ranges for most biomarkers in animal studies (differences are statistically based) plus a small effect in a key clinical endpoint can be potentially more adverse than a large effect in another clinical marker.

- C. Regarding non-cancer candidate PODs (Figure ES-4): Pg xxxviii (lines 14-16). Should explain the statement 'that BMDL modeling was largely unsuccessful due to data limitations' given the number of animal studies that are available. Maybe provide some general examples of major data limitations. Was this due the fact that the BMDL was at a much lower dose than the LOAEL or were there other reasons?
- D. Pg xxxvii, Lines 16-19 sentence needs clarification. It sounds like those studies that were eliminated for further analysis would have NOAELs available?? Am I misreading this?

Sec 4: General comments/questions related to section 4:

I noted on a Pubmed search for Dioxin, 16,000 publications came up. EPA's role, albeit difficult, is to pick the several most appropriate studies to establish an RfD on a scientific basis. With so many publications to choose from, even with schemes designed for best selections, determining 'the best' study is highly subjective. Thus, I would not want to second guess the 2 studies selected to establish the non-cancer PoD and I would recommend they be used. However, I think it's important that the document provide the reader more assurance that their selection was critically assessed and based on the best science. This may require that both the positive and negative aspects of the studies are fully discussed. A good argument is presented that the Seveso studies are creditable, consistent with LOAELs found for an array of effects in other animal studies and suitable for QRA. However, (and I use the Baccarelli Study Evaluation section as an example), additional points that might be included: 1 – *additional animal or epi supporting studies*: Although there are over 200 publications on thyroid effects and dioxin or dioxin-like compounds many of which are epi studies (according to PubMed), only Seo et. A., (1995) is discussed to help support or refute the observation from Baccarelli, and this study only had T4 levels and a LOAEL at least a log less sensitive. The NTP, 2006 chronic study appears to have been designed specifically for QRA and included thyroid hormone markers, including TSH. Why is this not discussed relative to the Baccarelli findings? There also appears to be a substantial number of Dx epi studies monitoring thyroid function. 2 – While I'm not an epidemiologist, there seems to be some design limitations in the Baccarelli study that should be mentioned. For example, there are many causes of elevated neonatal TSH including prematurity, severe illness, maternal medication, maternal thyroid disease and errors in the screening procedure. None of these were considered in the Baccarelli analysis. 3 – Can EPA provide data or better references to indicate that neonatal

elevated TSH in an adverse effect? Elevated thyroid-stimulating hormone (TSH) in newborns is certainly of concern and can lead to congenital hypothyroidism but without additional testing (e.g., later time points, inclusion of T3 and T4 levels, etc.) may have minimal significance. It would be helpful to know if there has been any follow-up in this population to determine if the elevation was transient or persistent. 4 - It appears from Table 5 and Fig 2 of the Baccarelli study that PCDDs, PCDFs, and coplanar PCBs were also associated with elevated TSH levels. This is mentioned in the EPA Study Evaluation (Pg 2-119) but not commented on – are the effects by Dx independent of other Dx-like compounds? 6 - It appears that the test population included individuals that resided in zone A at the time of the accident as well as those that moved in later. Does this complicate exposure reconstruction? 7. Are there no proposed MOA for Dx-induced thyroid effects? I believe in the case of Phenobarbital, elevated TSH is due to CYT P-450 activation which causes thyroid hormone metabolism. This may be of concern as EPA did not want to use Cyt P-450 activation as a critical target. If available, information from other studies on MOA might be useful.

Regarding the uncertainty associated with high dose acute exposure and its difference to average daily exposure: I don't have concerns with the model used to estimate daily exposure. However, it would be helpful to have some information assuring that the effects observed are not an acute/high dose phenomenon and would occur following low-dose chronic administration. Animal studies are replete with different toxic profiles following high-dose acute vs. low-dose chronic exposures. Are there any animal studies with Dx or Dx-like compounds comparing toxic endpoints between low-dose/accumulative exposure and acute/high dose that would help alleviate this concern? Since the Mocarrelli study grouped the males by age with the effected group ranging from 1 to 9 years of age at the time of exposure, it would be helpful to know if there exists any key time periods (windows) where a child would be more sensitive to effects.

Q 4.4: Taking a precautionary approach, I believe that in some instances biochemical changes, such as CYP induction and particularly oxidative stress can be used as a critical effect. For example, I would use it if it could be linked to a more acceptable clinical marker such as increased tissue 8-hydroxy-2'-deoxyguanosine (8-OHdG) DNA adducts as a precursor for cancer. However, for the Dx assessment their use would not be warranted. There are ample studies where there are better biomarkers of an adverse effect from Dx exposure that can be used.

Q 4.5: Yes. It appears appropriate as I understand it, presuming that the internal blood concentrations were adjusted for half-life. So for example, half-life adjusted for daily exposure vs. weekly exposure!

Q 4.6: Since there are other studies that determined TSH levels (albeit maybe not neonatal) and sperm quantity/quality, it would be useful to determine the POD for these markers when available (e.g., TSH in the NTP, 2006 study). The BMD modeling seems appropriate.

Q 4.7: Kinetic model output and application of UFs: The approach EPA used seems cautious but reasonable. I'm not a modeler but wouldn't this be taken into account since the confidence level was set to 95% and for dichotomous endpoints a BMR of 10% extra risk was used?

Q 4.8: The discussion of uncertainty is interesting. It may be my lack of familiarity with this issue but I had some difficulty separating hypothetical discussions vs. applicability to the Dx RfD and cancer slope factor. I wonder if it would be worthwhile to include an additional table that bullets key sources of uncertainty and provides a qualitative indicator (e.g., 1-5) of the relative uncertainty as it pertains specifically to the Dx assessment? Maybe it is already covered in Tables 6.1 and 6.2 but not easy to decipher.

General questions/clarification:

1. Assuming that the SAB finds no major problems with the document is this being considered the final external review or are there plans for another review or for the NAS to re-review?
2. Regarding the definition of an adverse effect. Obviously this would usually represents a biomarker that is indicative of a potential health effect. However, is the value based on a statistical difference, a value outside some normal reference range or at a value which would trigger medical treatment?

Luster Comments – Updated August 12, 2010

M.I. Luster Round 2 Comments 8-12-10:

First, let me repeat, but maybe more strongly than in my 1st round of comments, that, in general, the EPA document adequately responded to the key NAS recommendations, is clear and concise and there are no critical studies that are not addressed. My comments are intended to provide additional clarity to these issues (i.e., an attempt to extract every drop of juice from the orange). My major concerns were discussed, many at considerable length, at the July meeting, were included in the panel's summary comments, and the EPA

staff seemed to willing to address these concerns in the revisions. Thus, many of my comments will likely sound repetitious and hopefully will be addressed in the revised document.

- One additional issue came up during the discussions at the meeting. This relates to the rationale for not implementing TEFs in the assessment. This may have been discussed and adequately justified in the 2003 assessment, but obviously there were a number of public speakers as well as committee members who strongly disagree with the EPA decision not to have these included. TEFs were briefly discussed in the uncertainty analysis but this discussion provides a conflicting argument. I personally feel that their inclusion is scientifically justified and will be a useful endeavor as related to public health. However, if EPA believes inclusion will result in more confusion and or delay of completing the assessment, then I would agree that these not be part of the current assessment. In any case, the document should include a discussion justifying the decision.
- I would like to reinforce my concern regarding the discussions of uncertainty analysis. I mentioned in response to Q 4.8 that as a lab scientist and not a statistician, it appeared that there was a lot of theoretical discussion on how one can and difficulties in conducting an uncertainty analysis but limited information on specific TCDD issues. Maybe table 6.1 (NAS list of sources for uncertainty) or 6.2 (POD) can be revised or an additional table be added to provide better specificity to the effort.
- Section 2: Issues outstanding:
 - In most cases the rationale for the use of a study, the basis for the PoD selection, or model etc is justified on a clear scientific basis. In several instances, however, the basis for a decision is justified by the statement that it is 'consistent with EPA criteria or policy'. A sentence on the scientific basis of the policy is recommended.
 - Regarding study inclusion criteria (Fig ES-2 or Fig 2-3; pg 2-250): EPA indicated at the workshop that exclusion criteria Number 4 criteria 'accepted normal ranges' was inadvertently left in and should have been excluded from the final document - good. I still think some further discussion is required. The vast majority of studies reviewed would have qualified for the 1st and 2nd criteria since these criteria are normally required for most peer-reviewed publications. Thus, rejection of most of the studies was probably based upon exclusion criteria 3 (inconsistent toxicological practices). Are there some references that can be provided of what constitutes consistent toxicological practices or can this be better described?

- Issues remaining in section 4:
 - Although I believe they are adequate to establish a POD for establishing RfDs, I can't reiterate enough the need to better discuss the limitations and attributes of two Seveso studies. As with any retrospective, cross-sectional epidemiology study there is of course potential for misclassification, confounding, etc that EPA states. However, there are a number of other concerns that would help assure the reader that the studies were critically assessed and I will restate my earlier comments: *"A good argument is presented that the Seveso studies are credible, consistent with LOAELs found for an array of effects in other animal studies and suitable for QRA. However, (and I use the Baccarelli Study Evaluation section as an example), additional points that might be included: 1 – additional animal or epi supporting studies: Although there are over 200 publications on thyroid effects and dioxin or dioxin-like compounds many of which are epi studies (according to PubMed), only Seo et. al., (1995) is discussed to help support or refute the observation from Baccarelli, and this study only had T4 levels and a LOAEL at least a log less sensitive. The NTP, 2006 chronic study appears to have been designed specifically for QRA and included thyroid hormone markers, including TSH. Why is this (as well as other studies) not discussed relative to the Baccarelli findings? There also appears to be a substantial number of Dx epi studies monitoring thyroid function. 2 – While I'm not an epidemiologist, there seems to be some design limitations in the Baccarelli study that should be mentioned. For example, there are many causes of elevated neonatal TSH including prematurity, severe illness, maternal medication, maternal thyroid disease and errors in the screening procedure. None of these were considered in the Baccarelli analysis. 3 – Can EPA provide data or better references to indicate that neonatal elevated TSH in an adverse effect? Elevated thyroid-stimulating hormone (TSH) in newborns is certainly of concern and can lead to congenital hypothyroidism but without additional testing (e.g., later time points, inclusion of T3 and T4 levels, etc.) may have no or minimal significance. It would be helpful to know if there has been any follow-up in this population to determine if the elevation was transient or persistent. 4 - It appears from Table 5 and Fig 2 of the Baccarelli study that PCDDs, PCDFs, and coplanar PCBs were also associated with elevated TSH levels. This is mentioned in the EPA Study Evaluation (Pg 2-119) but not commented on – are the effects by Dx independent of other Dx-like compounds? 6 - It appears that the test population included individuals that resided in zone A at the time of the accident as well as those that moved in later. Does this complicate exposure reconstruction? 7. Are there no proposed MOA for Dx-induced thyroid effects? I believe in the case of Phenobarbital, elevated TSH is due to CYT P-450 activation which causes thyroid hormone metabolism. This may be of concern as EPA did not want to use Cyt P-450 activation as a critical target. If available, information from other studies on MOA might be useful."*

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- Because of the limitations in the 2 Seveso studies in terms of their use for risk assessment, there needs to be some discussion for not including a UF for database deficiency. Again, I think a more thorough discussion of the 2 Seveso studies would reduce any concerns of including an additional UF.

Dr. Paolo Mocarelli

Mocarelli Comments – September 1, 2010

After the July 13-15 Panel Meeting I have a few additional comments to make as follows:

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response

Analysis

The criteria used by EPA in their considerations to evaluate epidemiological and animal studies to select the key data set were transparent and clear.

Section 4. Reference Dose

I do not comment on Mocarelli et al (2008) and Bacarelli et al (2008) as they are papers coming from Seveso studies.

I only point out that the use of human data is more informative for populations than animal data even if these have given a great deal of information which can some time help to interpret the human one. Indeed, it must be remembered that in some cases, as in the case of TCDD toxicity for the liver, humans and animals differ greatly; in other cases, as for semen quality, the TCDD toxicity for rats and humans seems similar. Therefore, having the possibility to use human data can reduce the risk of interspecies difference in order to define a RfD.

Dr. Victoria Persky

Preliminary Comments – July 8, 2010

Section 2. Transparency and Clarity in Selection of Key Data Sets

Overall this section is quite responsive to NAS concerns. Criteria for the section of epidemiology and animal studies are clearly described and considerations of alternatives for exclusions discussed in detail and with clarity. I do wonder whether expanded discussion on suitability for inclusion of related articles on thyroid and diabetes would help the document? Material in the appendix suggests that the lack of an animal model for diabetes and paucity of published dose response data precludes its inclusion. Are there other reasons that this was not a primary endpoint? Thyroid homeostasis is complicated with effects in adults with occupational exposure not clearly understood. I think perhaps a more thorough discussion of these issues may be warranted if thyroid effects in newborns are to be highlighted. (Calvert 1999, Steenland 2001)

Section 5 Cancer Assessment

5.1 Weight of evidence: the weight-of-evidence is clearly described and appears scientifically justified.

5.2 Model of action: The discussions of liver, lung and thyroid cancer are clearly presented and the state of information on mechanism well addressed. I wonder, though, why there is not more discussion of lymphomas and associations with t(14:18) translocations in lymphocytes of dioxin-exposed individuals (Baccarelli et al 2006). Is it because of lack of data in animal models, difficulty in establishing dose effects or were there other reasons for exclusion?

5.3 The approach for selection of epidemiologic and animal bioassays clearly described and justified.

5.5 The Cheng paper (2006) seems well chosen. It extends previous analyses with dose dependent elimination, considers various lag periods and examines data without extremes as well as log transformed exposure values. There is discussion in the report of possible reasons for biases with inclusion of persons with extremely high exposures. I wonder whether some discussion of possible downregulation of AhR-dependent pathways with very high exposure might be a helpful addition as well (Landi 2003).

Overall, this is an excellent and thorough discussion of very complex issues and provides a sound rationale for the reanalysis and recommendations. My questions should be viewed as secondary to that general assessment and in no way detracting from a very complete and elucidating document.

1. Baccarelli A et al Carcinogenesis 27: 2001-2007, 2006
2. Calvert GM et al Occup Environ Med 56: 270-276, 1999
3. Cheng et al Risk Analysis 26: 1059-1071, 2006
4. Landi et al Carcinogenesis 24: 673-680, 2003
5. Steenland et al Occup Environ Med 2001: 58: 641-648, 2001

Persky Comments – Updated August 23, 2010

Section 2. Transparency and Clarity in Selection of Key Data Sets

Overall this section is quite responsive to NAS concerns. Criteria for inclusion of epidemiology and animal studies are clearly described and considerations of alternatives discussed in detail and with clarity. Criteria for exclusion of other studies, such as the Ranch Hand Study, are less clearly presented. The necessity of having low dose exposures seems reasonable. The criteria of TCDD being the sole exposure is more controversial and may exclude mixed exposures of TCDD with other dioxin related compounds on health endpoints for which there is emerging literature, such as diabetes (Steenland 2001). Similarly, there is limited discussion of dioxin effects on additional cancers, such as lymphoma (Bertazzi 2001). Inclusion of these endpoints might add support to the weight of evidence for the level of risk estimated from the thyroid study. Examination of neonatal thyroid effects in Seveso is of real interest. Effects in adults, however, are more complicated, with increased, rather than decreased free thyroxine index (FTI), apparent in some studies and the mode of action not clearly understood. (Calvert 1999)

Section 5 Cancer Assessment

5.1 Weight of evidence: the weight-of-evidence for included studies is clearly described and appears scientifically justified. There is less discussion regarding rare cancers, the power necessary to detect significant associations for rare events and the overall weight of evidence for dioxin's association with a wide variety of cancers, not just those included in this analysis.

5.2 Mode of action: The discussions of liver, lung and thyroid cancer are reasonably presented and the state of information on mechanism addressed. The report would be helped by some discussion of lymphomas and associations with t(14:18) translocations in lymphocytes of dioxin-exposed individuals (Baccarelli et al 2006). Less restrictive inclusion criteria, particularly regarding the necessity of primary TCDD exposure, would allow consideration of emerging data supporting lymphoma, as well as other rare cancers, in estimation of risk.

5.3 The approach for inclusion of epidemiologic studies and animal bioassays is clearly described and justified. There is less clarity around criteria for some of the excluded studies.

5.5 The Cheng paper (2006) seems well chosen. It extends previous analyses with dose dependent elimination, considers various lag periods and examines data without extremes as well as log transformed exposure values. The use of all-cancer mortality in this study focused on common cancers is reasonable. There is discussion in the report of possible reasons for biases with inclusion of persons with extremely high exposures. I wonder whether some discussion of possible down regulation of AhR-dependent pathways with very high exposure might be a helpful addition as well. (Landi 2003).

5.6 EPA has described major qualitative uncertainties. I must defer to others on the feasibility of developing quantitative estimates around these uncertainties.

5.7 The focus on TCDD exposures in isolation from exposures to dioxin-like compounds perhaps deserves further discussion, since in general TCDD exposures account for a small proportion of total TEQs and do not occur in isolation.

5.8 Justification for use of linear models is reasonable given the lack of data supporting non-linear low dose effects in existing animal and human studies.

Overall, this is a reasonable discussion of very complex issues that provides a sound rationale for the presented reanalysis and recommendations. Expanded discussion of inclusion of dioxin related compounds, weight of evidence for rare events, and integration of data supporting other endpoints would strengthen the report.

1. Baccarelli A et al Carcinogenesis 27: 2001-2007, 2006
2. Bertazzi PA et al American Journal of Epidemiology 153: 1031-44, 2001
3. Calvert GM et al Occup Environ Med 56: 270-276, 1999
4. Cheng et al Risk Analysis 26: 1059-1071, 2006
5. Landi et al Carcinogenesis 24: 673-680, 2003
6. Steenland et al Occup Environ Med 2001: 58: 641-648, 2001

Dr. Sandra Petersen

Preliminary Comments – July 8, 2010

EVALUATION OF EPA REANALYSIS OF KEY ISSUES RELATED TO DIOXIN TOXICITY AND RESPONSE TO NAS COMMENTS

Section 2 Transparency and clarity in the selection of key data sets for dose-response analyses

The process and criteria used by EPA to select key data for dose response analyses is clearly described in section 2.3 of this document and in the Executive Summary. In addition, the results of a literature search performed by EPA are available online, although clarity could be improved by providing search words used for the MedLine searches. A clear case for including high-quality human studies over animal studies is also made. However, the argument for having different criteria for epidemiological and animal bioassay studies could be stronger.

The statement, “The study criteria shown below and in Figure 2-3 for animal bioassay data reflect EPA’s preferences for TCDD-specific study inclusion, some of which are based on common practices and guidance for POD selection and RfD and OSF derivation” (p 2-5) does not help the reader understand the rationale for criteria. Please define what these common practices are more clearly. Concerns about criteria selection are below:

The requirement for peer-reviewed literature is reasonable, although journals with low impact factors generally have less stringent review processes. There is a clear and logical description of why a cutoff of 30 ng/kg-day was used (p 2.8 and 2.9) in the selection criteria. However, it is not clear why a specific statement of TCDD purity must be made explicitly. This is an important issue because TCDD is available from relatively few commercial sources and those sources certify purity of the chemical (typically \geq to 98% purity). Therefore, inclusion as one of the three major selection criteria seems somewhat arbitrary and the rationale could be clarified. Finally, the consideration that the “study design is consistent with standard toxicological practices” is unclear. It would be helpful to explain what aspects are different in toxicological studies than in physiological studies and the rationale for these differences.

Section 4 Reference Dose (Questions 4.4-4.8)

4.4 Was the decision to not use biochemical endpoints as potential critical effects scientifically justified and clearly described?

Overall, the discussion of this issue in the document was a bit repetitive, with no additional information provided in the various places where the issue is raised. A major weakness of the present assessment is that it does not adequately consider issues regarding identification of critical effects in the nervous system.

The brain is a lipid-rich organ and arguably one of the most important developmental targets of TCDD. Unfortunately, typical measures used to assess adverse effects in other organs (organ weight, observable lesions, blood markers of organ dysfunction or biochemical markers linked to disease) are inappropriate or unavailable for assessing neural impacts of TCDD exposure. Therefore, it is important to include biochemical endpoints such generation of ROS and activation of antioxidative enzymes, as well as changes in the number of serotonergic neurons, and

behaviors as critical endpoints in neurotoxicological assessments. I base these comments on information outlined below.

- 1) It is generally agreed that most, if not all, effects of dioxin are mediated by the AhR pathway. Expression of AhR, ARNT and ARNT2 is heterogeneous in the brain with only a few regions having high levels of expression (Petersen, Curran et al. 2000). This is very important because the brain is highly compartmentalized physiologically. Therefore, the brain cannot be examined without regard to anatomical region as one might examine other TCDD-sensitive tissues. For example, biopsies of liver from TCDD-exposed animals are likely to be representative of the entire organ, but that is not true of the brain. Only those regions that contain AhR are likely to respond and the responses among AhR-containing regions are likely to be different because different cell types have different co-activators and co-repressors as well as variable expression of steroid receptors with which AhR interact.

This concept is illustrated by the Hassoun *et al.* (Hassoun, Al-Ghafri et al. 2003) shows that developmental exposure to TCDD increases lipid peroxidation in the cerebral cortex and hippocampus, but not in the cerebellum and brainstem. In contrast, TCDD increases superoxide dismutase activity in the cerebellum and brainstem, but not in the cortex or hippocampus.

- 2) Disruption of cellular functions in the brain, particularly during development, can permanently alter a wide range of physiological processes, but effects are often unobservable until much later in life. This makes it difficult to connect exposures with physiological impairments that affect performance of the whole organism. Therefore, a logical way to evaluate TCDD adverse effects is to measure changes in biochemical markers known to alter neural function (which presumably constitutes an adverse effect) and to focus those studies on brain regions that contain AhR/ARNT or ARNT2.

The Hassoun study (Hassoun, Al-Ghafri et al. 2003) was omitted in the current analysis because changes in reactive oxygen species and antioxidant enzymes were not considered adverse effects. The brain is particularly susceptible to perturbations that increase oxidative stress because it has a high oxygen consumption rate, abundant unsaturated lipids and redox-capable metal ions, but relatively low levels of antioxidant enzymes. Moreover, the production of an imbalance between generation of reactive oxygen species and induction of antioxidant enzymes as demonstrated by Hassoun's group is deleterious. For example, development of neurons under oxidative stress conditions can suppress expression of mitochondrial manganese superoxide dismutase activity and increase cell death of mature neurons (Sompol, Ittarat et al. 2008). In addition, oxidative stress is linked to neurodegenerative disease processes that may start early in life (Kenneth, Nathan et al. 1995; Gu, Zhu et al. 2008). Thus, chronic or subchronic exposure to oxidative stress induced by TCDD in hippocampus and cortex (regions with relatively abundant AhR expression and key targets in Alzheimer's Disease and other neurodegenerative diseases that affect cognition) should be considered an adverse effect. The EPA states that , "it is impracticable to link the markers of oxidative stress to a toxicological outcome in the brain" (p 4-7), but this seems to eliminate some of the most sensitive markers of neurotoxicity. It would be helpful to clarify the scientific reasons for this practice. In addition, the EPA should clarify what endpoints are considered adverse in the nervous system.

- 3) It is particularly puzzling that 25-50% TCDD-dependent reductions in the numbers of serotonergic neurons in raphe nuclei was not be considered an adverse effect. The raphe nuclei

are the source of most of the serotonin in the brain and spinal cord. The permanent loss of these post-mitotic cells is clearly adverse considering that serotonin regulates such diverse functions as cognition, mood, neural control of gonadotropins and probably many functions not yet identified. Unfortunately, it would be very difficult to examine behavioral or other physiological manifestations of the loss of serotonin neurons because endpoints would depend on which serotonergic neurons were lost. Nevertheless, the findings described in this paper clearly show changes in brain structure (decreased number of cells and, consequently, raphe nuclei size) that should be considered an adverse effect. Therefore, this paper should be considered for further analysis.

- 4) Other than perhaps vasopressin and oxytocin release, there are no substances released by the brain into the bloodstream in response to disruptions in neural functions. Moreover, although disruption of cellular functions in the brain, particularly during development, can permanently alter a wide range of physiological processes, measurable endpoints are often unobservable until much later in life or are difficult to distinguish from downstream effects. For example, disruptions of the neural regulation of gonadotropin, thyrotropin or adrenocorticotropin release are difficult to separate from pituitary effects of TCDD.

On the contrary, learning, memory and motivation are valid outputs of brain function that have been assessed after TCDD exposure at relatively low doses. Therefore, it is not clear to me why the study of Markowski *et al.* (Markowski, Zareba et al. 2001) was not included in the literature considered for evaluation. In this study, investigators found a significant decrease in motivation at very low exposure levels.

The question of whether or not to consider CYP induction in the liver as an adverse effect is more complicated, because a key function of that organ is to metabolize compounds entering the blood stream through the gut. A concern that may not be germane to the present evaluation, but nevertheless important, is that CYP induction by AhR ligands has been linked to production of PCB metabolites that are more toxic than the parent compounds.

4.5 Was EPA's approach for averaging blood TCDD concentrations over the entire dosing period exposures and 24 h after last exposure valid in animal bioassay analysis (include discussion of intermittent and one day gestation exposure protocols)?

This is a difficult issue and the approach used by EPA was conservative. Considering the relatively long half-life it seems likely that there would be some accumulation over time. In addition, there is evidence that in rodents, TCDD is transferred by lactation up to 100 times more effectively than through placental transfer. Data from the Milbrath study (2009; [198044](#)) suggests that lactation substantially decreases half-life of TCDD in the mother. Thus, averaging without including a lactational component will likely underestimate doses in studies involving subchronic exposure. However, as I recall, there were few animal studies wherein this would be an issue. In multiple exposure protocols, the approach used by EPA (time-weighted average blood TCDD concentrations) seems reasonable.

For single gestational exposures, considering the peak blood TCDD concentration as the most relevant exposure metric may or may not be accurate. For example, in the case of sexual differentiation of the brain, a dose of TCDD given at GD8 would likely not have an effect on masculinization until 8-10 days later, so the relevant dose during the "critical period" would be less than the initial dose.

4.6 Were BMD modeling of animal bioassay data and choice of PODs from these analyses scientifically justified?

The description of modeling was clear and the assumptions seem valid. The decision to use models with higher BMDLs or AICs but much better fit to the lower response data seems to address NAS concerns. The only area for improvement I detected was a more thorough accounting of how many of the data sets were not modeled because no models satisfied the acceptance criteria. It was not clear how many of the data sets could be modeled without refitting by fixing the control mean or other methods.

4.7 Was the approach of using kinetic extrapolation for PODs prior to applying uncertainty factors scientifically justified and clearly described?

The approach was conservative and protective of human health. It was also clearly described. I have no alternative approaches to suggest.

4.8 Was EPA's qualitative discussion of uncertainty in the RfD well justified and clear? XXXXX working on this

Relevant (somewhat relevant?) to the issue of confidence based on concordance of data, a recent study reported not cited in this document might be important. In animals with gestational-onset hypothyroidism, there is impairment of sperm maturation in the epididymis in adulthood, as well as a significant decrease in the ratio of male:female pups delivered by untreated females mated to males who had been hypothyroid during gestation .

References:

- Gu, F., M. Zhu, et al. (2008). "Enhanced oxidative stress is an early event during development of Alzheimer-like pathologies in presenilin conditional knock-out mice." Neuroscience Letters 440(1): 44-48.**
- Hassoun, E. A., M. Al-Ghafri, et al. (2003). "The role of antioxidant enzymes in TCDD-induced oxidative stress in various brain regions of rats after subchronic exposure." Free Radical Biology and Medicine 35(9): 1028-1036.**
- Kenneth, H., H. Nathan, et al. (1995). "Brain Regional Correspondence Between Alzheimer's Disease Histopathology and Biomarkers of Protein Oxidation." Journal of Neurochemistry 65(5): 2146-2156.**
- Markowski, V. P., G. Zareba, et al. (2001). "Altered operant responding for motor reinforcement and the determination of benchmark doses following perinatal exposure to low-level 2,3,7,8-tetrachlorodibenzo-p-dioxin." Environ Health Perspect 109(6): 621-7.**
- Petersen, S. L., M. A. Curran, et al. (2000). "Distribution of mRNAs encoding the arylhydrocarbon receptor (AhR), arylhydrocarbon receptor nuclear translocator (ARNT) and ARNT2 in the rat brain and brain stem." J. Comp. Neurol. 427: 428-439.**

Sompol, P., W. Ittarat, et al. (2008). "A neuronal model of Alzheimer's disease: An insight into the mechanisms of oxidative stress-mediated mitochondrial injury." Neuroscience 153(1): 120-130.

Petersen Comments – Updated July 23, 2010

EVALUATION OF EPA REANALYSIS OF KEY ISSUES RELATED TO DIOXIN TOXICITY AND RESPONSE TO NAS COMMENTS

Section 2 Transparency and clarity in the selection of key data sets for dose-response analyses

The process and criteria used by EPA to select key data for dose response analyses is clearly described in section 2.3 of this document and in the Executive Summary. In addition, the results of a literature search performed by EPA are available online, although clarity could be improved by providing search words used for the MedLine searches. A clear case for including high-quality human studies over animal studies is also made. However, the argument for having different criteria for epidemiological and animal bioassay studies could be stronger.

The statement, “The study criteria shown below and in Figure 2-3 for animal bioassay data reflect EPA’s preferences for TCDD-specific study inclusion, some of which are based on common practices and guidance for POD selection and RfD and OSF derivation” (p 2-5) does not help the reader understand the rationale for criteria. Please define what these common practices are more clearly. Concerns about criteria selection are below:

The requirement for peer-reviewed literature is reasonable, although journals with low impact factors generally have less stringent review processes. There is a clear and logical description of why a cutoff of 30 ng/kg-day was used (p 2.8 and 2.9) in the selection criteria. However, it is not clear why a specific statement of TCDD purity must be made explicitly. This is an important issue because TCDD is available from relatively few commercial sources and those sources certify purity of the chemical (typically \geq to 98% purity). Therefore, inclusion as one of the three major selection criteria seems somewhat arbitrary and the rationale could be clarified. Finally, the consideration that the “study design is consistent with standard toxicological practices” is unclear. It would be helpful to explain what aspects are different in toxicological studies than in physiological studies and the rationale for these differences.

Section 4 Reference Dose (Questions 4.4-4.8)

4.4 Was the decision to not use biochemical endpoints as potential critical effects scientifically justified and clearly described?

Overall, the discussion of this issue in the document was a bit repetitive, with no additional information provided in the various places where the issue is raised. A major weakness of the present assessment is that it does not adequately consider issues regarding identification of critical effects in the nervous system.

The brain is a lipid-rich organ and arguably one of the most important developmental targets of TCDD. Unfortunately, typical measures used to assess adverse effects in other organs (organ

weight, observable lesions, blood markers of organ dysfunction or biochemical markers linked to disease) are inappropriate or unavailable for assessing neural impacts of TCDD exposure. Therefore, it is important to include biochemical endpoints such as generation of ROS and activation of antioxidative enzymes, as well as changes in the number of serotonergic neurons, and behaviors as critical endpoints in neurotoxicological assessments. I base these comments on information outlined below.

- 1) It is generally agreed that most, if not all, effects of dioxin are mediated by the AhR pathway. Expression of AhR, ARNT and ARNT2 is heterogeneous in the brain with only a few regions having high levels of expression (Petersen, Curran et al. 2000). This is very important because the brain is highly compartmentalized physiologically. Therefore, the brain cannot be examined without regard to anatomical region as one might examine other TCDD-sensitive tissues. For example, biopsies of liver from TCDD-exposed animals are likely to be representative of the entire organ, but that is not true of the brain. Only those regions that contain AhR are likely to respond and the responses among AhR-containing regions are likely to be different because different cell types have different co-activators and co-repressors as well as variable expression of steroid receptors with which AhR interact.

This concept is illustrated by the Hassoun *et al.* (Hassoun, Al-Ghafri et al. 2003) showing that developmental exposure to TCDD increases lipid peroxidation in the cerebral cortex and hippocampus, but not in the cerebellum and brainstem. In contrast, TCDD increases superoxide dismutase activity in the cerebellum and brainstem, but not in the cortex or hippocampus.

- 2) Disruption of cellular functions in the brain, particularly during development, can permanently alter a wide range of physiological processes, but effects are often unobservable until much later in life. This makes it difficult to connect exposures with physiological impairments that affect performance of the whole organism. Therefore, a logical way to evaluate TCDD adverse effects is to measure changes in biochemical markers known to alter neural function (which presumably constitutes an adverse effect) and to focus those studies on brain regions that contain AhR/ARNT or ARNT2.

The Hassoun study (Hassoun, Al-Ghafri et al. 2003) was omitted in the current analysis because changes in reactive oxygen species and antioxidant enzymes were not considered adverse effects. The brain is particularly susceptible to perturbations that increase oxidative stress because it has a high oxygen consumption rate, abundant unsaturated lipids and redox-capable metal ions, but relatively low levels of antioxidant enzymes. Moreover, the production of an imbalance between generation of reactive oxygen species and induction of antioxidant enzymes as demonstrated by Hassoun's group is deleterious. For example, development of neurons under oxidative stress conditions can suppress expression of mitochondrial manganese superoxide dismutase activity and increase cell death of mature neurons (Sompol, Ittarat et al. 2008). In addition, oxidative stress is linked to neurodegenerative disease processes that may start early in life (Kenneth, Nathan et al. 1995; Gu, Zhu et al. 2008). Thus, chronic or subchronic exposure to oxidative stress induced by TCDD in hippocampus and cortex (regions with relatively abundant AhR expression and key targets in Alzheimer's Disease and other neurodegenerative diseases that affect cognition) should be considered an adverse effect. The EPA states that, "it is impracticable to link the markers of oxidative stress to a toxicological outcome in the brain" (p 4-7), but this seems to eliminate some of the most sensitive markers of neurotoxicity. It would be helpful to clarify the scientific reasons for this practice. In addition, the EPA should clarify what endpoints are considered adverse in the nervous system.

- 3) It is particularly puzzling that 25-50% TCDD-dependent reductions in the numbers of serotonergic neurons in raphe nuclei was not considered an adverse effect. The raphe nuclei are the source of most of the serotonin in the brain and spinal cord. The permanent loss of these post-mitotic cells is clearly adverse considering that serotonin regulates such diverse functions as cognition, mood, neural control of gonadotropins and probably many functions not yet identified. Unfortunately, it would be very difficult to examine behavioral or other physiological manifestations of the loss of serotonin neurons because endpoints would depend on which serotonergic neurons were lost. Nevertheless, the findings described in this paper clearly show changes in brain structure (decreased number of cells and, consequently, raphe nuclei size) that should be considered an adverse effect. Therefore, this paper should be considered for further analysis.

The question of whether or not to consider CYP induction in the liver as an adverse effect is more complicated, because a key function of that organ is to metabolize compounds entering the blood stream through the gut. A concern that may not be germane to the present evaluation, but nevertheless important, is that CYP induction by AhR ligands has been linked to production of PCB metabolites that are more toxic than the parent compounds.

4.5 Was EPA's approach for averaging blood TCDD concentrations over the entire dosing period exposures and 24 h after last exposure valid in animal bioassay analysis (include discussion of intermittent and one day gestation exposure protocols)?

This is a difficult issue and the approach used by EPA was conservative. Considering the relatively long half-life it seems likely that there would be some accumulation over time. In addition, there is evidence that in rodents, TCDD is transferred by lactation up to 100 times more effectively than through placental transfer. Data from the Milbrath study (2009; [198044](#)) suggests that lactation substantially decreases half-life of TCDD in the mother. Thus, averaging without including a lactational component will likely underestimate doses in studies involving subchronic exposure. However, as I recall, there were few animal studies wherein this would be an issue. In multiple exposure protocols, the approach used by EPA (time-weighted average blood TCDD concentrations) seems reasonable.

For single gestational exposures, considering the peak blood TCDD concentration as the most relevant exposure metric may or may not be accurate. For example, in the case of sexual differentiation of the brain, a dose of TCDD given at GD8 would likely not have an effect on masculinization until 8-10 days later, so the relevant dose during the "critical period" would be less than the initial dose.

4.6 Were BMD modeling of animal bioassay data and choice of PODs from these analyses scientifically justified?

The description of modeling was clear and the assumptions seem valid. The decision to use models with higher BMDLs or AICs but much better fit to the lower response data seems to address NAS concerns. The only area for improvement I detected was a more thorough accounting of how many of the data sets were not modeled because no models satisfied the acceptance criteria. It was not clear how many of the data sets could be modeled without refitting by fixing the control mean or other methods.

I understand that the dose response to TCDD is probably not linear, but it would be exceedingly difficult to model, particularly for the nervous system. One would need to include not just TCDD-activation of AhR binding to DNA, but TCDD activation of membrane-associated AhR. In addition, AhR activation of gene expression depends on interactions with other nuclear receptors (i.e. steroid receptors). AhR activation by TCDD also regulates gene expression through "tethering" mechanisms that involve AP1 and SP1 sites. Finally activated AhR is a potent E3 ubiquitin ligase that regulates steroid receptors.

4.7 Was the approach of using kinetic extrapolation for PODs prior to applying uncertainty factors scientifically justified and clearly described?

The approach was conservative and protective of human health. It was also clearly described. I have no alternative approaches to suggest.

4.8 Was EPA's qualitative discussion of uncertainty in the RfD well justified and clear?

I found the discussion thoughtful and complete. Perhaps it could be strengthened including a discussion of why data from animal models and humans were quantitatively (but not qualitatively) different. For example, inclusion of data regarding inter-species differences in levels or affinity of AhR in various target organs, as well as differences in levels of steroid receptors with which AhR acts, could be included.

Relevant to the issue of confidence based on concordance of data between human and animal studies, a recent study not cited in the document may be important. Anbalagan *et al.* (Anbalagan, Sashi et al. 2010) found that rats with gestational-onset hypothyroidism had impaired epididymal sperm maturation. Moreover, there was significant a decrease in the ratio of male:female pups delivered by untreated females mated to males who had been hypothyroid during gestation. These findings are interesting in view of evidence provided in the Mocarelli and Baccarelli studies and other work showing altered sex ratios in progeny of TCDD exposed populations.

References:

- Anbalagan, J., A. M. Sashi, et al. (2010). "Mechanism underlying transient gestational-onset hypothyroidism-induced impairment of posttesticular sperm maturation in adult rats." Fertility and Sterility 93(8): 2491-2497.**
- Gu, F., M. Zhu, et al. (2008). "Enhanced oxidative stress is an early event during development of Alzheimer-like pathologies in presenilin conditional knock-out mice." Neuroscience Letters 440(1): 44-48.**
- Hassoun, E. A., M. Al-Ghafri, et al. (2003). "The role of antioxidant enzymes in TCDD-induced oxidative stress in various brain regions of rats after subchronic exposure." Free Radical Biology and Medicine 35(9): 1028-1036.**
- Kenneth, H., H. Nathan, et al. (1995). "Brain Regional Correspondence Between Alzheimer's Disease Histopathology and Biomarkers of Protein Oxidation." Journal of Neurochemistry 65(5): 2146-2156.**
- Petersen, S. L., M. A. Curran, et al. (2000). "Distribution of mRNAs encoding the arylhydrocarbon receptor (AhR), arylhydrocarbon receptor nuclear translocator (ARNT) and ARNT2 in the rat brain and brain stem." J. Comp. Neurol. 427: 428-439.**

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

Sompol, P., W. Ittarat, et al. (2008). "A neuronal model of Alzheimer's disease: An insight into the mechanisms of oxidative stress-mediated mitochondrial injury." Neuroscience 153(1): 120-130.

Dr. Karl Rozman

Preliminary Comments – July 8, 2010

General Charge Questions

1.1 Is the draft Response to Comments clear and logical? Has EPA objectively and clearly presented the three key NRC recommendations?

Yes and No. There are more than three recommendations (see pg. 26/27). There are arbitrary exclusion criteria for study selection. For example, abandoning the TEF concept excluded my hepta-bioassay, which has good dose-response data.

1.2 Are there other critical studies that would make a significant impact on the conclusions of the hazard characterization or dose-response assessment of the chronic noncancer and cancer health effects of TCDD?

Yes, many including many of my own studies.

Specific Charge Questions

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis.

2.1 Is this Section responsive to the NAS concern about transparency and clarity in data-set selection for dose-response analysis?

Yes, but it is excessively verbose and the criteria are not appropriate from the toxicological point of view.

2.2 Are the epidemiology and animal bioassay study criteria/considerations scientifically justified and clearly described?

They are clearly described but scientifically not justified.

2.3 Has EPA applied the epidemiology and animal bioassay study criteria/considerations in a scientifically sound manner? If not, please identify and provide a rationale for alternative approaches.

EPA applied its own study selection criteria which are scientifically not sound. It is particularly troubling that previous efforts (TEF, Ah receptor, etc.) of EPA are now not part of the criteria.

Section 3. The Use of Toxicokinetics in the Dose-Response Modeling for Cancer and Noncancer Endpoints.

3.1 The 2003 Reassessment utilized first-order body burden as the dose metric. In the draft Response to Comments document, EPA used a physiologically-based pharmacokinetic (PBPK) model (Emond et al., 2004, 2005, 2006) with whole blood concentration as the dose metric rather than first-order body burden. This PBPK model was chosen, in part, because it includes a biological description of the dose-dependent elimination rate of TCDD. EPA made specific modifications to the published model based on more recent data. Although lipid-adjusted serum concentrations (LASC) for TCDD are commonly used as a dose metric in the literature, EPA chose whole blood TCDD concentrations as the relevant dose metric because serum and serum lipid are not true compartments in the Emond PBPK models (LASC is a side calculation proportional to blood concentration).

There is nothing wrong with PBPK analysis, although it is not needed when you have human data.

Please comment on:

3.1.a. The justification of applying a PBPK model with whole blood TCDD concentration as a surrogate for tissue TCDD exposure in lieu of using first-order body burden for the dose-response assessment of TCDD.

3.1.b. The scientific justification for using the Emond et al. model as opposed to other available TCDD kinetic models.

3.1.c. The modifications implemented by EPA to the published Emond et al. model.

3.1.d. Whether EPA adequately characterized the uncertainty in the kinetic models.

3.4. Several of the critical studies for both noncancer and cancer dose-response assessment were conducted in mice. A mouse PBPK model was developed from an existing rat model in order to estimate TCDD concentrations in mouse tissues, including whole blood.

Please comment on:

3.2.a. The scientific rationale for the development of EPA's mouse model based on the published rat model (Emond et al., 2004, 2005, 2006).

3.2.b. The performance of the mouse model in reference to the available data.

3.2.c. Whether EPA adequately characterized the uncertainty in the mouse and rat kinetic models. Please comment specifically on the scientific justification of the kinetic extrapolation factor from rodents to humans.

3.3 Please comment on the use of Emond et al. PBPK model to estimate human intakes based on internal exposure measures.

3.4 Please comment on the sensitivity analysis of the kinetic modeling (see Section 3.3.5).

3.5 Both EPA's noncancer and cancer dose-response assessments are based on a lifetime average daily dose. Did EPA appropriately estimate lifetime average daily dose? If not, please suggest alternative approaches that could be readily developed based on existing data.

Section 4. Reference Dose

4.1. The Mocarelli et al. (2008) and Baccarelli et al. (2008) studies were selected as co-critical studies for the derivation of the RfD. Is the rationale for this selection scientifically justified and clearly described? Please identify and provide the rationale for any other studies that should be selected, including the rationale for why the study would be considered a superior candidate for the derivation of the RfD. In addition, male reproductive effects and changed in neonatal thyroid hormone levels, respectively, were selected as the co-critical effects for the RfD. Please comment on whether the selection of these critical effects is scientifically justified and clearly described. Please identify and provide the rationale for any other endpoints that should be selected as the critical effect.

Are these the best scientific data to establish RfD?

4.2 In the Seveso cohort, the pattern of exposure to TCDD is different from the average daily exposure experienced by the general population. The explosion in Seveso created a high dose pulse of TCDD followed by a low level background dietary exposure in the exposed population. In the population, this high dose pulse of TCDD was slowly eliminated from body tissues over time. There is uncertainty regarding the influence of the high-dose pulse exposure on the effects observed later in life.

Not much uncertainty.

4.2.a. Mocarelli et al. (2008), reported male reproductive effects observed later in life for boys exposed to the high dose pulse of TCDD between the ages of 1 and 10. EPA identified a 10 year critical exposure window. In the development of the candidate RfD, EPA used an exposure averaging approach that differs from the typical approach utilized for animal bioassays. EPA determined that the relevant exposure should be calculated as the mean of the pulse exposure and the 10-year critical exposure window average. Please comment on the following:

Why 1 to 10 and not 1 to 8 or 1 to 12. It should have been pre- and post-puberty.

4.2.a.i. EPA's approach for identifying the exposure window and calculating average exposure for this study.

4.2.a.ii. EPA's designation of a 20% decrease in sperm count (and an 11% decrease in sperm motility) as a LOAEL for Mocarelli et al. (2008).

4.2.b. For Baccarelli et al. (2008), the critical exposure window occurs long after the high-dose pulse exposure. Therefore, the variability in the exposure over the critical exposure window is likely to be less than the variability in the Mocarelli et al. subjects. EPA concluded that the reported maternal exposures from the regression model developed by Baccarelli et al. provide an appropriate estimate of the relevant effective dose as opposed to extrapolating from the measured infant TCDD concentrations to maternal exposures. Additionally, EPA selected a LOAEL of 5 μ -units TSH per ml blood in neonates; as this was established by World Health Organization (WHO) as a level above which there was concern about abnormal thyroid development later in life. Please comment on the following:

4.2.b.i. EPA's decision to use the reported maternal levels and the appropriateness of this exposure estimate for the Baccarelli et al. study.

4.2.b.ii. EPA's designation of 5- μ units TSH per ml blood as a LOAEL for Baccarelli et al. (2008).

4.3 Please comment on the rationale for the selection of the uncertainty factors (UFs) for the RfD. If changes to the selected UFs are proposed, please identify and provide a rationale.

4.4 EPA did not consider biochemical endpoints (such as CYP induction, oxidative stress, etc.) as potential critical effects for derivation of the RfD for TCDD due to the uncertainties in the qualitative determination of adversity associated with such endpoints and quantitative determination of appropriate response levels for these types of endpoints in relation to TCDD exposure. Please comment on whether this decision is scientifically justified and clearly described.

I don't see lowered uncertainty with the selected data since there are no dose-responses but there are clear reproducible dose-responses for CYP etc. Moreover there are plenty of toxic dose-responses in other animal studies.

4.5 In using the animal bioassays, EPA averaged internal blood TCDD concentrations over the entire dosing period, including the 24 hours following the last exposure. Please comment on EPA's approach for averaging exposures including intermittent and one day gestation exposure protocols.

4.6 Please comment on the benchmark dose (BMD) modeling conducted by EPA to analyze the animal bioassay data and EPA's choice of points of departure (PODs) from these studies.

4.7 For the animal bioassay modeling, EPA applied the kinetic extrapolation at the level of the POD prior to applying the uncertainty factors because EPA has less confidence in the kinetic model output at lower doses reflective of the RfD. Please comment on whether this approach was scientifically justified and clearly described.

4.8 Please comment as to whether EPA's qualitative discussion of uncertainty in the RfD is justified and clearly described.

4.5, 4.6, 4.7, 4.8 – What is the point if available data are not used for RfD nor for cancer?

Section 5. Cancer Assessment

- a. Weight of Evidence Cancer Descriptor: The 2003 Reassessment concluded that TCDD is a "known human carcinogen." In the current draft Response to Comments document, EPA concluded that under the 2005 Guidelines for Carcinogen Risk Assessment (U.S. EPA, 2005) TCDD is "carcinogenic to humans." Is the weight-of-evidence characterization scientifically justified and clearly described?**

Looking at animal data implies weight of evidence analysis but it is not used in this document because everything is based on epidemiology. This is a trivial statement. All compounds are carcinogenic to humans if another effect does not kill the individual before carcinogenicity can become manifest (other than physical limitations).

- b. Mode of Action: The mode of action of a carcinogen can inform identification of hazards and approaches used for a dose-response assessment. The mode of carcinogenic action for TCDD has not been elucidated for any tumor type. EPA concluded that, while interaction with the Ah receptor is likely to be a necessary early event in TCDD carcinogenicity in experimental animals, the downstream events involved are unknown.**

We know about mode of action for carcinogenicity of TCDD as much or more as for any other known carcinogen.

5.2.a. Are the available data related to mode(s) of action for the carcinogenicity of TCDD appropriately characterized and clearly presented?

5.2.b. Do the available data support EPA's conclusion that the overall mode(s) of action for TCDD-induced carcinogenesis is largely unknown? Please comment on whether this evaluation is clearly described.

Data were selected to support a policy decision.

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- c. Is EPA's approach for selecting data sets from the key epidemiologic studies and animal bioassays identified for cancer dose response modeling scientifically justified and clearly described?**

The studies selected do not prove that point.

5.4 For the animal bioassay data, potential cancer oral slope factors (OSFs) were calculated by linear extrapolation (using a linear, non threshold cancer approach) from the point of departure (POD). EPA also estimated the composite risk of the occurrence of several tumor types from the animal cancer bioassay data.

5.4.a. Please comment on whether the approach for estimating cancer risk, including the use of tumor modeling of the TCDD animal cancer bioassay data, is scientifically justified and clearly described.

5.4.b. Please comment on the choice of using a BMDL01 as the POD for the development of candidate oral slope factors derived from the TCDD animal cancer bioassays.

5.5 EPA selected Cheng et al. (2006) – an analysis of the NIOSH occupational cohort – as the critical study for oral slope factor (OSF) development. This study was chosen because it considers dose-dependent elimination of TCDD rather than first-order kinetics.

A small improvement, probably not worth the effort.

5.5.a. Please comment on whether the rationale for this selection is scientifically justified and clearly described. Please identify and provide the rationale for any other studies that should be considered and provide a critical evaluation of the study and of its suitability for meeting the goals of a quantitative cancer assessment.

Rozman et al. in Food Chem. Toxicol.

5.5.b. Cheng et al. (2006) analyzed all-cancer mortality. Please comment on the use of all-cancer mortality as the basis of the OSF.

Problematic. Very weak data made questionable by lack of site specificity.

5.5.c. Please comment on whether the use of the Emond PBPK model in the estimation of risk-specific doses from the Cheng et al. dose-response modeling results is scientifically justified and clearly described.

See my comment on PBPK.

5.5.d. EPA elected to use the log linear relationship of fat concentration and rate ratio to estimate risk-specific doses at all risk levels. EPA could have estimated a POD for cancer risk itself at a single risk level (BMR) for extrapolation to the origin. Please comment on EPA's choice of extrapolation approach.

5.5.e. The slope factor derived from Cheng et al. (2006) was extrapolated below the background TCDD exposure levels experienced by the NIOSH cohort. Please comment on this extrapolation.

Linear prediction of DDT, dieldrin, aflatoxin → 153,000 liver cancers and there are only 7,000 per year. What is a prediction worth that does not predict realistically?

5.6 Please comment on whether EPA has clearly described the major qualitative uncertainties in the derivation of the OSF.

5.7 EPA did not consider dioxin-like compounds (DLCs) in the cancer dose-response modeling because the occupational exposures in the available cohorts were primarily to TCDD. Background DLC exposures were not incorporated in the dose-response modeling because EPA judged that it was not possible to disaggregate the responses from background exposure to DLCs and occupational exposure to TCDD. Please comment on whether this approach is scientifically justified and clearly described.

Not particularly convincing after EPA spent 20+ years to establish the validity of the TEF concept. There is scientific validity in TEFs, see previous EPA documents and also my studies.

5.8 The NRC suggested that EPA consider nonlinear approaches for the assessment of TCDD carcinogenicity. In the Response to Comments, EPA presents two illustrative nonlinear approaches for cancer, but considers both inappropriate to use because lack of MOA information.

Here, EPA is clearly non-responsive to NRC.

5.8.a. Please comment on these two illustrative nonlinear approaches including EPA's conclusions regarding the limitations of these approaches.

Unconvincing.

5.8.b. Are there other nonlinear approaches that could be readily developed based on existing data for the assessment of TCDD carcinogenicity? If so, please suggest alternative approaches and describe their utility and suitability for meeting the goals of a quantitative cancer assessment.

Hepta-lung cancer (Rozman et al. 2005)

Rozman comments

page 76

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment

This is not my area of expertise.

6.1 Please comment on the discussion in this Section. Is the response clearly presented and scientifically justified?

6.2 Please comment on EPA's overall conclusion that a comprehensive quantitative uncertainty analysis is not feasible.

6.2.a. Please comment on the discussion in Section 6 regarding volitional uncertainty and how this type of uncertainty limits the ability to conduct a quantitative uncertainty analysis.

6.3 Throughout the document (including the Appendices), EPA presents a number of limited sensitivity analyses (e.g., toxicokinetic modeling, RfD ranges, cancer OSF ranges, cancer RfD development). Please comment on the approaches used, and the utility of these sensitivity analyses in clarifying potential significant uncertainties.

This document would be improved with a discussion on reversibility issues.

Rozman Comments – Updated September 1, 2010

According to my notes, the EPA-Document in general is non-responsive to the NAS recommendations. The NAS-Committee identified not 3 but 7 key findings to be addressed by EPA. There was no recommendation to abandon the DLC concept which EPA and other scientific and regulatory bodies established over the past 20+ years as a scientifically plausible and experimentally justified principle (TEF/TEQ). Nor was there any recommendation to use human epidemiology data only to derive an RfD. The EPA-Document's criteria for this purpose were unscientific. Truncating the dose-response at $\leq 1 \mu\text{g}/\text{kg}/\text{day}$ for cancer or $\leq 30 \text{ mg}/\text{kg}/\text{day}$ for non-cancer, if I remember correctly, makes the uncertainty about these studies extremely large. There are plenty of reliable animal data from which an RfD could be derived in accordance with the NAS recommendation.

The cancer risk assessment is particularly troubling. Instead of taking the recommendation of NAS seriously...” to compare cancer risk using both a linear model and a nonlinear model consistent with receptor-mediated mechanism of action...” the authors explain voluminously why this cannot or should not be done. All of this is done in the face of an enormous uncertainty whether or not dioxins cause cancer in humans at doses that cause chloracne, particularly since the original analysis of the Fingerhut Study yielded negative results in terms of total cancer incidence. Instead of chasing molecules, a world-wide chloracne registry should have been established a long time ago to monitor cancer incidence in people with unequivocal chloracne (a total of thousands of people vs. e.g. office workers) and the cancer question could have been or still could be answered unequivocally. The only part of the document that is acceptable from the scientific point of view is the PBPK modeling, which is just a minor point in the NAS recommendation. In my view PBPK

modeling is not needed to support dioxin risk assessment. Actually PBPK modeling needs the dioxin data for its improvement. Again the EPA-Document defied the recommendation of NAS: "EPA should continue to use body burden as the preferred dose metric..." I could go on and on with criticism but I will not do so because I do not know how much more work EPA is willing to apply to this project. Hopefully, my criticism will be taken for what it is, a constructive attempt to separate science from preconceived policy decisions and thereby to help EPA to create positions that are scientifically defensible and to avoid positions which on the long-run would lead to problems worse than the agency is currently facing.

Regarding dose-response modeling there is extensive data on Ah receptor dose-responses. One example can be found with many earlier citations in Sloop and Lucier (1988). This is a relatively old citation; there is a large number of more recent ones.

Dr. Arnold Schechter

Preliminary Comments – July 9, 2010

Initial thoughts regarding EPA's Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments
Arnold Schechter

In keeping with our new schedule, this will be a preliminary overview of my review of the document.

The EPA document is a thoughtful, carefully prepared response to the NAS

I believe this was also the case for the earlier EPA dioxin document prepared in 1995. A glossary including definitions would be extremely useful. Some terms are defined in the document; however, having easily accessible definitions would be very useful for persons who are not expert in many of the scientific aspects of this complex, multidisciplinary document.

having ready access to the complete set of articles cited would be valuable and save the time currently required to look needed citations up. It appears an attempt was made to allow such access though the system seems imperfect as I have been unable to open the links.

The goal of this document was to explore the potential health effect of 2,3,7,8-TCDD specifically. However, the data used includes not only 2,3,7,8-TCDD but also PCDDs, PCDFs, and DL PCBs. While it is appropriate to include additional compounds, to more accurately assess their relation to 2,3,7,8-TCDD, I recommend using the toxic equivalent (TEQ)/order of magnitude approach. This approach is the best method currently available for estimating dioxin toxicity. It has repeatedly been validated since its inception during the Binghamton State Office Building electrical transformer incident in the early 1980s. While not perfect, it appears to be the best approach that exists. The 5 year WHO revisions appear to be a reasonable approach to updating this model and providing the best current order of magnitude approach to estimating dioxin toxicity

It will be difficult and I believe pointless for EPA to repeatedly update their Dioxin Reassessment document if many changes are requested and numerous attempts are made to satisfy these requests. It might be more useful to prepare a document, as EPA has done several times, and publish it not as a draft but as a final document for that time period. Periodic updates would be appropriate, similar to updates provided by the IOM/NAS with "Veterans and Agent Orange," ATSDR's Toxicological Profiles, and the WHO for dioxin-like TEFs. It appears to me that at present EPA adequately describes the scientific state of affairs, including various options for evaluating the incomplete data, which exists now and which will always be less than desired.

The linear no threshold model used by EPA has been debated for many years and it appears no resolution will be made at this time. Simply listing reasons for and against using this approach rather than revising the dioxin document repeatedly seems most appropriate. It will be important

to describe the differences between these approaches for low dose risk estimates, including how much difference there is for various approaches.

Written comments from the Panel might be considerably more useful for the Science Advisory Board if all public comments were first completed and then expert opinions of the Panel presented in writing. Each panelist could comment on all points and chapters where indicated rather than selecting some panelists to focus on a given chapter or two. This is especially true since the review process and public meetings have now increased far beyond the original time frame. The opinions presented in person would be more valuable than only reading the EPA document.

The articles cited seem somewhat selective compared to the more comprehensive reviews of the literature in IARC dioxin document No. 69 on dioxins, the Veterans and Agent Orange NAS series published every two years, and the periodic ATSDR Toxicological Profiles of dioxins, dibenzofurans, and PCBs. Extensive literature reviews can also be found in "Dioxins and Health," 2nd Ed., Schecter and Gasiewicz, 2003.

References sometimes cite government documents rather than the seminal articles. This should be discouraged even though it is common these days. Seminal articles should usually be cited. References sometimes are of abstracts only, e.g., Thiess et al., rather than peer reviewed articles. I would suggest placing these in a separate category. They should, however, only be used if they are the sole documents presenting the data.

The IOM/NAS series, "Veterans and Agent Orange," has been updated every two years since its first publication in 1994. If the 2000 publication is cited, it would be reasonable to also cite later editions.

Below are publications which I feel may be useful but which have not been cited in the EPA document. These publications from the USA include human findings that address partitioning of various organs and tissues, the effects of cooking, and findings in vegans with lower levels of dioxins and other fat soluble persistent organic pollutants (POPs) which may act in a similar fashion to dioxins and dioxin like chemicals.

- The Yusho and Yucheng incidents are useful in addressing the issue of TEFs and human ill health from dioxins, dibenzofurans, and other dioxin like compounds. They are also useful in documenting shorter half lives of elimination at higher levels.
- Inclusion of the Baughman and Meselson article in EHP from the 1970s documents the change from chloracne to dioxin tissue measurement in humans using human milk from Vietnam and the USA and food from Vietnam might be instructive in following the change in methods of dioxin exposure assessment. The Seveso finding that chloracne is seen only at high levels of TCDD in blood and not in all with high levels is an important point.

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- Articles summarizing detection of elevated TCDD and also TBDD 30-40 years after exposure in workers and a chemist might be useful in discussion of estimating exposure over time.
- The decrease of almost 70% of levels of dioxins in a mother nursing twins for up to 2 and a half years illustrates two things: depuration from nursing and one method of eliminating dioxins from the body.

I was struck by the modest mention of actual measured dioxin levels in Americans when such levels exist and can be used for partitioning purposes. This is striking to me when a Japanese paper is cited for partitioning (Maruyama 2002) rather than a number of American papers, such as the autopsy paper of Schecter and Mes for PCBs or the Ryan and Schecter paper for dioxins and dibenzofurans. The Japanese paper does not appear to list actual data but instead provides derived or calculated values. In addition, I am also concerned as most dioxins and DL compounds enter humans by dietary route in animal fat. The Japanese diets traditionally differ from American diets by having less meat and more fish and vegetables. When US data exists, it should be considered.

It was not apparent in my review of the document that fetal and early childhood dioxin levels were considered to a sufficient extent

Levels measured in food seem not to be emphasized in this document as much as I would have expected.

Consideration of special populations such as vegans where dioxin, dibenzofuran, and PCB levels have been measured could be useful in providing estimates of low dose exposure for a portion of the population. This has been done for dioxins and also for PBDEs which are believed to partition in a similar fashion. Partitioning ratios also exist in blood, milk, and some other organs in humans. I favor beginning an exercise with measured data from humans, then animal studies, and finally modeling approaches. This would provide a more solid basis for coming to conclusions. Measured levels of dioxins in human semen compared to blood exist and may be useful in developmental studies where dioxin transfer from semen to egg may, as speculated by Hatch and Stein, possibly play a role in male mediated adverse effects on the fetus.

Human partitioning data from Patterson and Needham of CDC, Peter Fuerst of Germany, and Schecter and colleagues for components of blood, blood and milk, and various organs of the body exist and should be used when applicable to increase the data used for EPA's reports and conclusions. These data may also be used to note apparent contradictions, such as the question of why similar dioxin levels are sometimes found whether whole blood or serum is measured.

Selected potential citations from work done with my colleagues and that may be of value in refining the EPA document are listed below:

- Baughman R, Meselson M. 1973. An analytical method for detecting TCDD (dioxin): levels of TCDD in samples from Vietnam. *Environmental Health Perspectives* 5: 27-35.

- Ryan J, Schechter A, Lizotte R, Sun W-F, Miller L. 1985. Tissue Distribution of Dioxins and Furans in Humans from the General Population. *Chemosphere* 14(6): 929-932.
- Eadon G, Kaminsky L, Silkworth J, Aldous K, Hilker D, Okeefe P, et al. 1986. Calculation of 2,3,7,8-Tcdd Equivalent Concentrations of Complex Environmental Contaminant Mixtures. *Environmental Health Perspectives* 70: 221-227.
- Schechter A, Dekin A, Weerasinghe NCA, Arghestani S, Gross ML. 1988. Sources of Dioxins in the Environment - a Study of Pcds and Pcdfs in Ancient, Frozen Eskimo Tissue. *Chemosphere* 17(4): 627-631.
- Schechter A, Ryan JJ. 1988. Polychlorinated Dibenzo-Para-Dioxin and Dibenzofuran Levels in Human Adipose Tissues from Workers 32 Years after Occupational Exposure to 2,3,7,8-Tcdd. *Chemosphere* 17(5): 915-920.
- Schechter A, Mes J, Davies D. 1989. Polychlorinated Biphenyl (Pcb), Ddt, Dde and Hexachlorobenzene (Hcb) and Pcd/F Isomer Levels in Various Organs in Autopsy Tissue from North-American Patients. *Chemosphere* 18(1-6): 811-818.
- Schechter A, Papke O, Ball M. 1990. Evidence for Transplacental Transfer of Dioxins from Mother to Fetus: Chlorinated Dioxin and Dibenzofuran Levels in the Livers of Stillborn Infants. *Chemosphere* 21(8): 1017-1022.
- Schechter A, Ryan JJ, Constable JD, Baughman R, Bangert J, Furst P, et al. 1990. Partitioning of 2,3,7,8-Chlorinated Dibenzo-Para-Dioxins and Dibenzofurans between Adipose-Tissue and Plasma-Lipid of 20 Massachusetts Vietnam Veterans. *Chemosphere* 20(7-9): 951-958.
- Schechter A, Papke O, Ball M, Ryan J. 1991. Partitioning of Dioxins and Dibenzofurans: Whole Blood, Blood Plasma, and Adipose Tissue. *Chemosphere* 23(11): 1913-1919.
- Schechter A, Ryan J, Constable J. 1992. Partitioning of Dioxin and Dibenzofuran Congeners Between Plasma and Cell Fractions of Blood From 10 Adult Male Patients. *Chemosphere* 25(12): 2017-2022.
- Schechter A, Startin J, Wright C, Kelly M, Papke O, Lis A, et al. 1994. Congener-specific Levels of Dioxins and Dibenzofurans in U.S. Food and Estimated Daily Dioxin Toxic Equivalent Intake. *Environmental Health Perspectives* 102: 962-966.
- Masuda Y, Haraguchi K, Kuroki H, Ryan J. 1995. Change of PCDF and PCB concentrations in the blood of Yucheng and Yusho patients for 25 years. *Fukuoka Igaku Zasshi* 86(5): 178-183.
- Schechter A, Dai LC, Le TBT, Quynh HT, Minh DQ, Cau HD, et al. 1995. Agent-Orange and the Vietnamese - the Persistence of Elevated Dioxin Levels in Human Tissues. *American Journal of Public Health* 85(4): 516-522.
- Schechter A, McGee H, Stanley JS, Boggess K, BrandtRauf P. 1996a. Dioxins and dioxin-like chemicals in blood and semen of American Vietnam veterans from the state of Michigan. *American Journal of Industrial Medicine* 30(6): 647-654.
- Schechter A, Papke O, Lis A, Ball M, Ryan JJ, Olson JR, et al. 1996b. Decrease in milk and blood dioxin levels over two years in a mother nursing twins: Estimates of decreased maternal and increased infant dioxin body burden from nursing. *Chemosphere* 32(3): 543-549.
- Schechter A, Startin J, Wright C, Papke O, Ball M, Lis A. 1996c. Concentrations of polychlorinated dibenzo-p-dioxins and dibenzofurans in human placental and fetal tissues

from the US and in placentas from Yu-Cheng exposed mothers. *Chemosphere* 32(3): 551-557.

- Schechter A, McGee H, Stanley JS, Boggess K, BrandtRauf P. 1997. Dioxins and dioxin-like chemicals in blood and semen of American Vietnam veterans from the state of Michigan (vol 30, pg 647, 1996). *American Journal of Industrial Medicine* 31(3): 370-371.
- Schechter A, Olson JR. 1997. Cancer risk assessment using blood dioxin levels and daily dietary TEQ intake in general populations of industrial and non-industrial countries. *Chemosphere* 34(5-7): 1569-1577.
- Schechter A, Dellarco M, Papke O, Olson J. 1998a. A comparison of dioxins, dibenzofurans and coplanar PCBs in uncooked and broiled ground beef, catfish and bacon. *Chemosphere* 37(9-12): 1723-1730.
- Schechter A, Kassis I, Papke O. 1998b. Partitioning of dioxins, dibenzofurans, and coplanar PCBs in blood, milk, adipose tissue, placenta and cord blood from five American women. *Chemosphere* 37(9-12): 1817-1823.
- Schechter A, Ryan JJ, Papke O. 1998c. Decrease in levels and body burden of dioxins, dibenzofurans, PCBs, DDE, and HCB in blood and milk in a mother nursing twins over a thirty-eight month period. *Chemosphere* 37(9-12): 1807-1816.
- Schechter A, Cramer P, Boggess K, Stanley J, Papke O, Olson J, et al. 2001a. Intake of dioxins and related compounds from food in the US population. *Journal of Toxicology and Environmental Health-Part A* 63(1): 1-18.
- Schechter A, Dai LC, Papke O, Prange J, Constable JD, Matsuda M, et al. 2001b. Recent dioxin contamination from Agent Orange in residents of a southern Vietnam city. *Journal of Occupational and Environmental Medicine* 43(5): 435-443.
- Schechter A, Quynh HT, Pavuk M, Papke O, Malisch R, Constable JD. 2003. Food as a source of dioxin exposure in the residents of Bien Hoa City, Vietnam. *Journal of Occupational and Environmental Medicine* 45(8): 781-788.
- Schechter A, Paepke O, Harris TR, Chi Tung K. 2006a. Partitioning of polybrominated diphenyl ether (PBDE) congeners in human blood and milk. *Toxicological & Environmental Chemistry* 88(2): 319-324.
- Schechter A, Paepke O, Tung KC, Brown T, Musumba A. 2006b. Changes in polybrominated diphenyl ether (PBDE) levels in cooked food. *Toxicological & Environmental Chemistry* 88(2): 207-211.
- Schechter A, Papke O, Harris TR, Tung KC, Musumba A, Olson J, et al. 2006c. Polybrominated diphenyl ether (PBDE) levels in an expanded market basket survey of US food and estimated PBDE dietary intake by age and sex. *Environmental Health Perspectives* 114(10): 1515-1520

Schechter Comments – Updated August 24, 2010

This preliminary report follows the second meeting of the dioxin review panel in Washington DC several points were made by EPA SAB administration: the task with which we are charged is to respond to the NAS remarks on EPA documents regarding the dioxin reassessment. Some of our discussion in Washington involved scientific issues in

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

the original dioxin reassessment rather than the National Academy of Sciences critique and EPA's response to NAS-which was not what we were charged with reviewing.

Many of us in Washington mentioned that our expertise was limited to specific areas of science and that none of us, to the best of my knowledge felt competent to evaluate all scientific areas covered by EPA and NAS in the reports we were sent to evaluate.

My area of expertise lies primarily in medicine, public health, exposure assessment and epidemiology, mostly of dioxins and dioxin like chemicals including some other POPs and endocrine disrupting compounds.

I am writing this preliminary report following the Washington meeting primarily to discuss broad issues. I note with interest that the National Academy of Sciences publishes documents every two years entitled "Veterans and Agent Orange". These summarize the literature on health effects of dioxins and the herbicides contained in Agent Orange and related chemicals used in Vietnam. These are prepared and published by a panel assisted by staff of the Institute of Medicine of the National Academy of Sciences every two years based on an evaluation of the published peer-reviewed literature. The EPA began its dioxin reassessment over a decade ago but its release has been slowed down repeatedly, most recently by having the National Academy of Sciences write a report to which EPA must respond. The NAS in turn, does not have similar critiques and responses required to prior to publishing its similar or at least closely related document every two years.

I believe the EPA dioxin reassessments including updates and response to NAS have been of good scientific quality. They are neither comprehensive nor perfect. It has often been said that perfection is the enemy of the good. I believe we are seeing this happening by the many delays in release of EPA's dioxin reassessment. We, like IOM/NAS are seeing new scientific literature on dioxins published yearly in large amounts which show the need to publish and then update EPA's dioxin document in order to remain current. At this point in time EPA's response to the NAS review of their most recent dioxin reassessment seem adequate to me. This most recent update of EPA's dioxin reassessment is years old and continues the process of its reevaluating dioxin which began over a decade ago.

I believe EPA has prepared a reasonable dioxin reassessment and although the NAS has made interesting comments EPA has replied in a reasonable fashion. Not all scientists can agree with all of EPA's dioxin reassessment nor with all of EPA's response to NAS. However, I see no point in further holding up release of the EPA dioxin document.

As an example, whether low-dose dioxin biological response is linear or nonlinear or whether or not there is a threshold for dioxin effects or pathology will continue to be debated by qualified scientists for a considerable time. How to evaluate the published peer-reviewed scientific literature on various aspects of rodent and human dioxin levels and biological, clinical and other outcomes can and will be debated for many years to come by qualified scientists and risk assessors.

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

In this report, I have not gone into individual scientific details, which I plan to do in a subsequent report in the near future, I believe that EPA has responded in a reasonable fashion to NAS critique of their document based to a certain extent on discussion at the recent Washington DC meeting. I believe EPA staff at that meeting agreed to further refine the EPA response in a manner that I believe most but not all of us at the panel meeting found to be reasonable. Therefore, for these reasons and in light of the use to which the EPA dioxin reassessment will be used in the United States and worldwide, I believe it best that the updated EPA response to be accepted and officially published before this year is over. This should of course include comments made by panel members and the public and conclusions agreed upon by the majority of those at the most recent Washington DC meeting of our panel.

Dr. Allen Silverstone

Preliminary Comments – July 8, 2010

Initial comments, July 8th, 2010 to the Draft document “EPA’s Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments (“Response to Comments”).

Despite a huge amount of very interesting and significant new work in this draft, I think we all have to be mindful that (as this draft report shows), there is a continuing production of relevant information including new aspects of how TCDD and related compounds induce pathophysiological changes. I am pleased that this reanalysis even includes data published in 2009, but it is likely that new data, and possibly dramatically new implications could become available in the future.

Therefore, I strongly believe we must do the best with what we have now, and commit to a regular reappraisal just as is done with the Vietnam Veteran’s Agent Orange Project (I think they are actually on a biennial review but it could be longer).

What follows are some very preliminary personal thoughts on this draft report, which I presume we will develop over the next 3 months with public input and discussions.

General Charge questions:

1.1. The draft Response to the NRC questions, is clear and coherent. As a member of that NRC panel, I think EPA has summarized the major recommendations, and attempted to address them.

1.2 At this time I don’t know of other critical studies that would impact on this hazard characterization and dose-response assessment. There is more recent work in development that is directed at measuring AhR activation in complex mixtures and exposures, but it is unclear whether this can be subsumed by TEQ concepts, or whether there are truly novel interactions with AhR that lead to different effects, and that might have implications for lower dose end points than this report characterizes.

2 Transparency and Clarity in the Selection of Key Data Sets for Dose-Response analysis:

2.1 Overall this Section is responsive to the NAS concern about transparency and clarity in data-set selection for dose response analysis. Not only are overarching criteria presented, but there is specific discussion of key data sets (including some previously used in this Health Effects Analysis) describing their shortcomings and justification for exclusion or even inclusion.

There seems to be a varying acceptance of DLCs (planar PCB’s, for example) in determining relevant data, but I think the report overstates how “pure” its analysis is of dioxin/TCDD in its data selection. The few times I believe I saw inclusion of DLCs into key subsets, there was also evidence of pure dioxin involvement. However, both in approaching this problem, with the growing awareness of “endo-dioxins”, and non-dioxin like compounds reacting with the AhR, risk analysis for this standard should also consider incorporation of these non 2,3,7,8 TCDD endpoints (I don’t think anyone would argue with chlorinated furans being included). It seems reasonable to this reviewer to exclude such compounds as ortho-substituted PCB’s at this time,

not only because their mechanism of action may be very different, but because the reasons why some of these provoke such distinctly different responses is only now being examined (e.g., it is clear that certain natural products including tryptophan metabolites, like FICZ, or bilirubin metabolites like ITE trigger development of an inflammatory immune response while the Dioxin like PCB's, seem to promote a suppressive, T-regulatory cell phenomenon.

The discussion of whether certain end points that are measured represent adverse health effects and their elimination because the authors reject this is in some cases, arguable. For example, exclusion of the Sugita-Konishi, 2003 study (where low dose TCDD in a perinatal animal study led to reductions in spleen size, increases in CD4 T-cells, and decreased effectiveness in response to challenge by *Listeria*) is excluded because the linkage between TCDD and immune function in this case is not demonstrated or clear. Similarly, the 2009 ANL-EPA meeting recommended using two other immune studies (one by De Vito, the other from Seveso) seems to have been ignored because a reduction in IgG is not considered an adverse health effect (the same for thymic atrophy, since the thymus normally atrophies in development). I therefore would suggest that there is more exclusion that is necessary, and I do not think, especially for non cancer endpoints, this is a good thing. At the same time, the LOAEL/NOAEL numbers in these come within the same range as determined by the selected studies. In previous iterations, the Dutch study with immune and neurodevelopment endpoints in a perinatal cohort was used much more than this document does.

However, this is not an issue of insufficient description, but the question is whether these studies should be excluded.

The choice of thyroid effects as a major sensitive endpoint could also be argued against, or supplemented by consideration of diabetes (including in the Seveso cohorts), or immunopathology determined end points. This should be talked about.

Thyroid homeostasis is a difficult endpoint to consider, for choosing a regulatory level. The most serious consequences of abnormal thyroid circuit activity/production probably happens in utero, and while indication of abnormal levels can be done in Italy due to a public health program that takes heel sticks for thyroid at the start of life. It is unclear that measurements of TSH/FT4 etc. are clearly understood as adverse effects.

Some further review of this rejection of diabetes and assorted immunological endpoints vs. utilizing thyroid endpoints might be in order. I do not think that inclusion will affect the BMD/RfD determinations by more than an order of magnitude (and probably much less), but the stronger the weight of evidence for the numbers derived, the better for acceptance.

Section 3: Toxicokinetics. The development of EPA's mouse model based on the the published Emond, et. al., model seems reasonable, in integrating the adverse effects in mice to the adverse effects in rats. My initial review of the performance of this modified model is favorable, but I would need to examine this further.

Section 4 Development of a reference dose.

4.1 See above. Using the Mocarelli (2008) and Baccarelli (2008) as co-critical endpoints for RfD while excluding the IgG reduction in Seveso cohort, and diabetes (again in Seveso) because there is no animal model (yet) concerns this reviewer. As at a first pass, as similar RfD would be derived, it might be good to add more weight to the evidence for the RfD.

4.4 While the reasons for exclusion of biochemical endpoints because of lack of actual adverse health effects linked to those end points is explained, I do not agree with excluding data derived from such endpoints. As more, and more data is developed using such mechanistic precursors of damage (like altered gene expression... e.g., recent work demonstrating low dose PCB exposure inducing genes in animal models associated with ADHD in humans argues that an exclusion of some of these endpoints is probably overly restrictive and exclusionary. Similarly reductions in serotonin due to TCDD exposure could have connection with many psychiatric problems. Lots of depression treatment uses manipulation of serotonin uptake to improve the state of disease.

4.5 The rationale for averaging internal blood TCDD concentrations over an entire dosing period might be a simplifying assumption that doesn't allow for either circadian alterations (and since data is coming out that AhR plays some role in controlling hemopoietic stem cell development, or that establish ranges of exposure within the same study (i.e., adverse effects actually only caused in a narrow time window when levels are maximum).

4.6 I am struck by the range of PODs and modeling from so many studies all being within an order of magnitude of an average value. As to the question of modeling and justification in the approach and uncertainty discussion I will need to do further work to see if I have any disagreements with the RfD/BMD-10 calculations. The approach seems to be justified but again may be too restrictive, or failing to allow for a range of values covering confidence limits or uncertainty.

Dr. Mitchell Small

Preliminary Comments – July 8, 2010

Initial Review of

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment

Mitchell Small
July 7, 2010

Charge Questions

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment

6.1. Please comment on the discussion in this Section. Is the response clearly presented and scientifically justified?

6.2. Please comment on EPA's overall conclusion that a comprehensive quantitative uncertainty analysis is not feasible.

6.2.a. Please comment on the discussion in Section 6 regarding volitional uncertainty and how this type of uncertainty limits the ability to conduct a quantitative uncertainty analysis.

6.3. Throughout the document (including the Appendices), EPA presents a number of limited sensitivity analyses (e.g., toxicokinetic modeling, RfD ranges, cancer OSF ranges, cancer RfD development). Please comment on the approaches used, and the utility of these sensitivity analyses in clarifying potential significant uncertainties.

Findings

The report addresses a broad range of philosophical and methodological issues in conducting an uncertainty analysis for TCDD toxicity, specifically for estimates of cancer oral slope factors and noncancer reference doses. The Section is successful in identifying the challenges involved in assessing uncertainty in toxicity estimates based on:

- A small set of available models for toxicokinetics, dose-response relationships, and low dose extrapolation, with limited application, testing, and verification; and

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

- A small set of animal bioassay, epidemiological or clinical/case studies, many with differing endpoints, dose metrics, and (in the case of the human studies) uncertain exposure and subject data.

As such, the Section provides many useful insights for EPA's Reassessment. However, in its discussion of available methods, the report is somewhat biased in its treatment of certain statistical methods which could address some of these issues (though it does note their potential contribution at the end of the Section, as part of ongoing or future studies) and overly pessimistic regarding our ability provide improved quantitative estimate for certain portions of the toxicity assessment. This is unfortunate since, in other Sections of the Reassessment, the report provides a very credible discussion of the range of scientific uncertainty in current knowledge regarding TCDD toxicokinetics and toxicity.

Methods that should be given a more extensive and balanced discussion, including more citations to the literature include:

Bayesian Hierarchical Modeling (for combing information from multiple studies):

Axelrad DA, Bellinger DC, Ryan LM, Woodruff TJ. Dose–response relationship of prenatal mercury exposure and IQ: an integrative analysis of epidemiologic data. *Environ Health Perspect* 2007;115:609–615.

Choi, T., M. J. Schervish, K. A. Schmitt and M. J. Small. 2010. Bayesian hierarchical analysis for multiple health endpoints in a toxicity study. *Journal of Agricultural, Biological, and Environmental Statistics*. Available online at: <http://www.springerlink.com/content/2h416p2581210773/fulltext.pdf>

Coull B., Menzetti M. and Ryan L. (2003) A Bayesian hierarchical model for risk assessment of methylmercury, *Journal of Agricultural, Biological and Environmental Statistics*, 8, 3, 253–270.

Ryan L. Combining data from multiple sources, with applications to environmental risk assessment. *Stat Med* 2008: 27(5): 698–710.

Bayesian Model Averaging (for considering more than one dose-response equation, allowing the data to weight their *relative* likelihood and contribution to the estimate):

Morales, Knashawn H., Joseph G. Ibrahim, Chien-Jen Chen, and Louise M. Ryan. 2006. “Bayesian Model Averaging With Applications to Benchmark Dose Estimation for Arsenic in Drinking Water.” *Journal of the American Statistical Association* 101 (473): 9–17.

These preliminary comments are from individual members of the SAB Dioxin Review Panel and do not represent consensus SAB advice or EPA policy. DO NOT CITE OR QUOTE. Updated as of October 12, 2010.

Viallefont, V., Raftery, A.E. and Richardson, S. (2001) Variable selection and Bayesian model averaging in case-control studies. *Statistics in Medicine* 20: 3215-3230.

Wheeler MW, Bailer AJ (2007). Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Risk Estimation." *Risk Analysis*, 27, 659-670.

Wheeler, M. W., Bailer, A. J. (2009). Comparing model averaging with other model selection strategies for benchmark dose estimation. *Environmental and Ecological Statistics* , **16** (1): 37–51.

Note: These Bayesian methods should not be referred to as “exotic”. For example, in agreeing with the Section 6 authors that these methods should be pursued in ongoing and future case studies, White et al. (2009) refer to them as “advanced”, rather than exotic. Specifically, they recommend that health scientists should:

Explore statistical approaches to model selection
Improvements to statistical approaches for model selection, such as model averaging, should be pursued. Case study applications of these advanced statistical approaches will identify potential strengths and weaknesses of these approaches and their significance for risk characterization.

White et al. (2009)

R.H. White, I. Cote, L. Zeise, M. Fox, F. Dominici, T.A. Burke, P.D. White, D. Hattis, J.M. Samet, State-of-the-science workshop report: issues and approaches in low dose–response extrapolation for environmental health risk assessment, *Environ. Health Perspect.* 117 (2009) 283–287.

Distributional (Probability Tree) Methods for considering alternative assumptions and models at various stages of the toxicity assessment.

These methods do rely upon expert judgment, but can provide a basis for ongoing integration and value of information assessment as new studies and knowledge accumulate over time (Brusick et. al., 2008). As described in Small (2008):

The distributional approach for characterizing uncertainty in cancer risk assessment was developed by Evans, Sielken, and co-workers beginning in the 1990s^(2–10) and has also been referred to as information analysis, weight-of-evidence analysis, the comprehensive methodology, and comprehensive realism.^(8–10) The method has since been acknowledged in a number of reviews of cancer risk assessment practice and research needs,^(11–13) and applied in various forms for risk assessment of different chemical compounds.^(14–19)

The motivation for the distributional approach is the recognition that the use of a single set of assumptions for the components of a cancer risk assessment, whether default, conservative, or otherwise, fails to capture the full range of plausible or likely relationships, how these relationships depend upon our current state of knowledge, the implications for computed values of potency or unit risk, and the opportunities for improved estimates. The distributional approach thereby enables consideration of a "portfolio-of-mechanisms" that may contribute to carcinogenesis.⁽²⁰⁾

- 2. Holland, C. D., Sielken, R. L. Jr. (1993). Quantitative Cancer Modeling and Risk Assessment . (Chapter 7). Englewood Cliffs , NJ : Prentice Hall.
- 3. Evans, J. S., Graham, J. D., Gray, G. M., Sielken, R. L. Jr. (1994). A distributional approach to characterizing low-dose cancer risk. *Risk Analysis* , 14 (1), 25–34.
- 4. Evans, J. S., Graham, J. D., Gray, G. M., Sielken, R. L. Jr. (1995). A distributional approach to characterizing low-dose cancer risk. In S. Olin, W. Farland, C. Park, L. Rhomberg, R. Scheuplein, T. Starr, J. Wilson (Eds.), *Low-Dose Extrapolation of Cancer Risks* (pp. 253–274). Washington , DC : ILSI Press.
- 5. Sielken, R. L. Jr. (1993). Evaluation of chloroform risk to humans. *The Toxicology Forum*, 1993 Annual Winter Meeting, February 15–17, 1993, The Capitol Hilton, Washington, DC.
- 6. Evans, J. S., Gray, G. M., Sielken, R. L., Jr., Smith, A. E., Valdez-Flores, C., Graham, J. D. (1994). Use of probabilistic expert judgment in distributional analysis of carcinogenic potency. *Regulatory Toxicology and Pharmacology* , 20 (1), 15–36.
- 7. Sielken, R. L., Jr., Valdez-Flores, C. (1999). Probabilistic risk assessment's use of trees and distributions to reflect uncertainty and variability and to overcome the limitations of default assumptions. *Special Issue of Environmental International on Modeling and Simulation* , 25 , 755–772.
- 8. Sielken, R. L. Jr. (1990). A weight-of-evidence approach to quantitative cancer risk assessment: Information analysis. In G. Schettler, D. Schmähl, T. Klenner (Eds.), *Risk Assessment in Chemical Carcinogenesis* . New York : Springer-Verlag. Proceedings of the Satellite Symposium on Risk Assessment in Chemical Carcinogenesis, Heidelberg, Germany August 24–25, 1990.
- 9. Sielken, R. L., Jr., Bretzlaff, R. S., Stevenson, D. E. (1995). Challenges to default assumptions stimulate comprehensive realism as a new tier in quantitative cancer risk assessment. *Regulatory Toxicology and Pharmacology* , 21 , 270–280.
- 10. Sielken, R. L., Jr., Valdez-Flores, C. (1996). Comprehensive realism's weight-of-evidence based distributional dose-response characterization. *Special Issue of the Human and Ecological Risk Assessment on Theoretical, Toxicological and Biostatistical*

Foundations for Deriving Probability Distribution Functions for Reference Doses and Benchmark Doses with Application to Carcinogens and Noncarcinogens , 2 (1), 175–193.

- 11. Boyce, C. P. (1998). Comparison of approaches for developing distributions for carcinogenic slope factors. *Human and Ecological Risk Assessment* , 4 (2), 527–577.
- 12. Moschandreas, D. J., Karuchit, S. (2002). Scenario-model-parameter—A new method of cumulative risk uncertainty analysis. *Environment International* , 28 (4), 247–261.
- 13. Zeise, L., Hattis, D., Andersen, M., Bailer, A. J., Bayard, S., Chen, C., Clewell, H., Conolly, R., Crump, K., Dunson, D., Finkel, A., Haber, L., Jarabek, A. M., Kodell, R., Krewski, D., Thomas, D., Thorslund, T., Wassell, J. (2002). Improving risk assessment: Research opportunities in dose response modeling to improve risk assessment. *Human and Ecological Risk Assessment* , 8 (6), 1421–1444.
- 14. Humphreys, S. H., Carrington, C., Bolger, M. (2001). A quantitative risk assessment for fumonisins B1 and B2 in US corn. *Food Additives and Contaminants* , 18 (3), 211–220.
- 15. Rai, S. N., Bartlett, S., Krewski, D., Paterson, J. (2002). The use of probabilistic risk assessment in establishing drinking water quality objectives. *Human and Ecological Risk Assessment* , 8 (3), 493–509.
- 16. Kirman, C. R., Sweeney, L. M., Teta, M. J., Sielken, R. L., Valdez-Flores, C., Albertini, R. J., Gargas, M. L. (2004). Addressing nonlinearity in the exposure-response relationship for a genotoxic carcinogen: Cancer potency estimates for ethylene oxide. *Risk Analysis* , 24 (5), 1165–1183.
- 17. Starr, T. B., Matanoski, G., Anders, M. W., Andersen, M. E. (2006). Workshop overview: Reassessment of the cancer risk of dichloromethane in humans. *Toxicological Sciences* , 91 (1), 20–28.
- 18. David, R. M., Clewell, H. J., Gentry, P. R., Covington, T. R., Morgott, D. A., Marino, D. J. (2006). Revised assessment of cancer risk to dichloromethane II. Application of probabilistic methods to cancer risk determinations. *Regulatory Toxicology Pharmacology* , 45 (1), 55–65.
- 19. Crump, K. S. (1994). Risk of benzene-induced leukemia: A sensitivity analysis of the Pliofilm cohort with additional follow-up and new exposure estimates. *Journal Toxicology and Environmental Health* , 42 , 219–242.
- 20. Cox, L. A. (2006). Quantifying potential health impacts of cadmium in cigarettes on smoker risk of lung cancer: A portfolio-of-mechanisms approach. *Risk Analysis* , 26 (6), 1581–1599.

Brusick, D., Small, M. J., Cavalieri, E. L., Chakravarti, D., Ding, X., Longfellow, D. G., Nakamura, J., Rogan, E. C., Swenberg, J. A. (2008). Possible genotoxic modes of action for naphthalene. *Regulatory Toxicology and Pharmacology*, 51 (2), 43–50.

Small, M.J. 2008. Methods for assessing uncertainty in fundamental assumptions and associated models for cancer risk assessment. *Risk Analysis*, 28(5): 1289-1307.

Specific Comments:

(To be expanded upon later)

1. Page 6-2. Add NRC(1996)
2. 6-3, bottom: margins → marginal
3. 6-4, line 9: the tone is too pedagogical (“This is not the place . . .”)
4. 6-5, I consider epistemic to mean unknown and aleatoric to mean inherently variable. So when (for example) body weight varies across a population, but with a distribution that is unknown, this exhibits both aleatoric and epistemic uncertainty.
5. 6-7, more examples of use of expert judgment for health assessment are available and should be cited.
6. 6-9, line 18, provide citations for dependence modeling
7. 6-10, line 4, add mention of methods that identify uncertain assumptions/parameters that are *important* – for determining whether the model is consistent with observed data (Hornberger and Spear) and for affecting a decision that is made as a result of the model (Merz et al.)
8. 6-16, line 20. Perhaps we can say that variability (and uncertainty) in the factors that are used to determine a particular UF can be considered in choosing the particular value of the UF.
9. 6-17, lines 3-14. I disagree with this assertion. This problem can be addressed using a Bayesian analysis with a beta conjugate for the uncertain response probability, p , with informationless (uniform) prior for p . The probability that “an experiment with a null response might have yielded a positive response” can be estimated from the predictive distribution (which will depend on the number of test animals in the original study that yielded zero responses) for the next experiment (with any number of exposed animals). I will bring an example to the meeting.

A Bayesian beta-binomial model for estimating “the probability that an experiment with a null response might have yielded a positive response”

M.J. Small
July 8, 2010

A beta distribution is a common model for representing uncertainty in an event occurrence probability, p (e.g., a failure rate for a machine part, or the probability of response for a targeted health outcome in a test animal tested with a given exposure). In a Bayesian analysis of binomial experiments, the beta serves as the conjugate distribution for p . That is, if the prior distribution for p is beta, so is the posterior distribution. In particular, if the prior distribution for p is $\text{Beta}(a_0, b_0)$, and a binomial experiment is conducted yielding y positives (in toxicity studies, y is the number of animals exhibiting a response) out of n trials (typically $n \sim 50$ test animals in a toxicity study), then the posterior distribution for p is $\text{Beta}(a = a_0 + y, b = b_0 + n - y)$. It is common in this type of analysis to assume a flat (or “informationless”) prior, achieved in this case by setting $a_0 = 1$ and $b_0 = 1$. This is equivalent to a uniform distribution over the range zero to one (the uniform distribution is a special case of the beta distribution), indicating that no value of p is a priori favored over any other. (Other priors could be considered if information is available to suggest that a particular range of p is more likely than others, but the flat prior is the simplest for this illustration.)

The conjugate beta distribution described above represents the epistemic uncertainty in the response rate, p , at a given exposure. However, even if p were known with certainty ($= p^*$), the number of positive responses in a test with n_n new test animals would exhibit inherent variability (aleatory uncertainty), given by a binomial distribution with parameters n_n and p^* . For the case where p is uncertain, the uncertainty in the number of positive responses, y_n in the next n_n trials includes *both* the uncertainty in p and the variability in y_n for a given p . This distribution is referred to as a *predictive distribution*. When the uncertainty in p is described by a beta distribution (as above) with parameters a and b , then the predictive distribution (the probability mass function) for y_n has a closed-form solution, referred to as the beta-binomial model:

$$p_{y_n}(y_n | n_n, a, b) = \binom{n_n}{y_n} \frac{B(a+y_n, b+n-y)}{B(a, b)}; \quad y_n = 0, 1, 2, \dots, n$$

where $\binom{n_n}{y_n}$ is the binomial coefficient,

$$B(a,b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$$

and $\Gamma()$ denotes the gamma function.

Figure A illustrates results of the beta-binomial model for the case discussed in Section 6 (page 6-17) where the initial study (at a particular dose rate) yielded $y = 0$ positive responses out of $n = 50$ trials. Figure A1 shows the posterior cumulative distribution function of p (beta with parameters $a = 1 + 0 = 1$, and $b = 1 + 50 - 0 = 51$). As indicated, the posterior estimate for p is now heavily skewed towards zero – and though while it seems that the range of the posterior distribution for p is limited to approximately $0 - 0.1$, some amount of probability density still remains over the entire range from zero to one.

Figure A2 shows the predictive distribution of y_n (beta binomial with $n_n = 50$, $a = 1$, $b = 51$). As indicated, given what was learned from the first experiment, the most likely outcome in the second experiment with $n_n = 50$ test animals is $y_n = 0$, however, this probability, $p(y_n = 0)$, is only 0.505. That is, there is nearly a 50 percent probability that there will be one or more positive responses in the next test of 50, given no positive responses in the first test (given the flat, uniform prior for p). This is a solution to the problem posed on lines 13 and 14 of page 6-17: finding “the probability that an experiment with a null response might have yielded a positive response,” achievable through the use of standard Bayesian methods.

An interesting question to pose is, how big must the initial n be with $y = 0$ to be $X\%$ sure that a second test with $n_n = 50$ will result in $y_n = 0$? The beta-binomial model allows this to be computed. For example, to be 80% confident ($X = 80$) that the next 50 animals tested will yield a null response, $n = 199$ initial animals must test null in the initial experimentation. To be 90% sure, $n = 449$ animals must test null in the initial trials at the targeted dose. A spreadsheet is provided for these calculations.

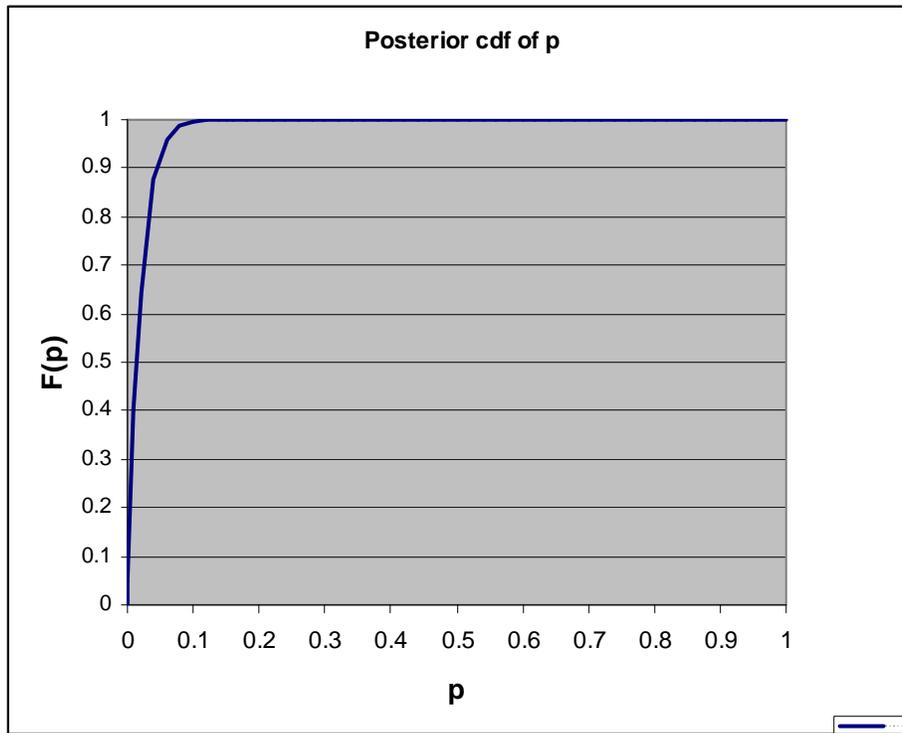


Figure A1 Posterior cdf of p

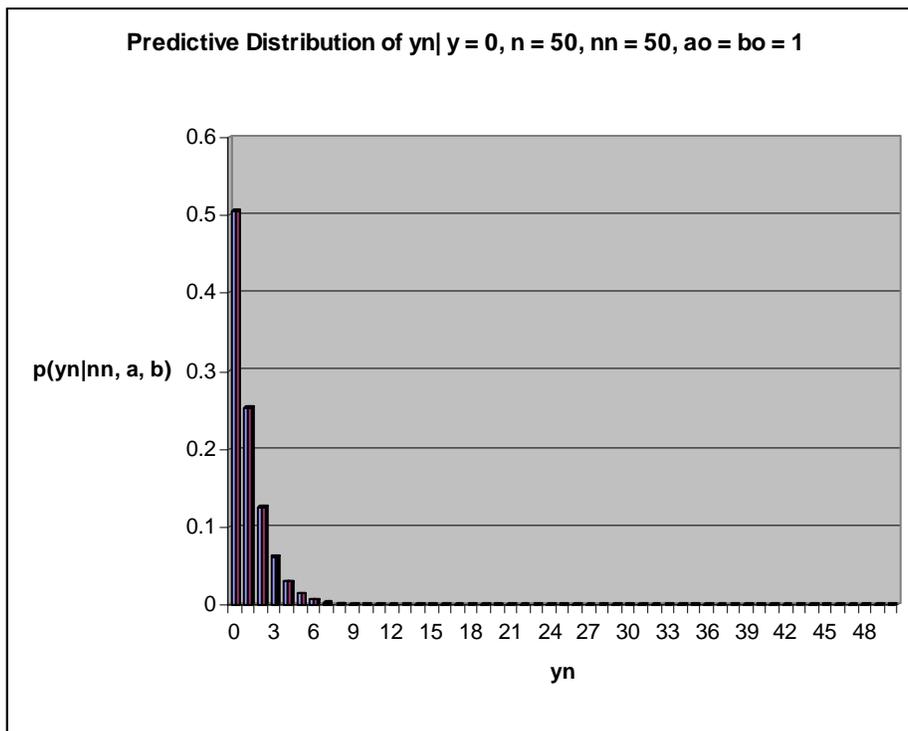


Figure A2. Predictive distribution of y_n given a second experiment with $n_n = 50$

Dr. Anne Sweeney

Preliminary Comments – July 9, 2010

Submitted by Anne Sweeney

General Comments: I would like to extend my gratitude to the EPA researchers for their impressive efforts in responding to the NAS concerns. Of particular note is the active inclusion of the public in both the identification of additional sources of dioxin-related materials and in participation in the February workshop and in the current Dioxin Panel Review activities.

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis

2.1. Is this section responsive to the NAS concern about transparency and clarity in data-set selection for dose-response analysis?

I believe that the EPA has sufficiently addressed these concerns. The EPA's collaboration with Argonne National Laboratory and invitation to the public to engage in updating the literature search to identify all appropriate studies for evaluation, as well as the conduct of the Dioxin Workshop in February of 2009, were instrumental in enhancing the transparency and clarity regarding the process of selection of studies for the dose-response analysis. The development of clear criteria for study evaluation and inclusion (discussed below) were crucial in resolving the concerns raised by the NAS.

2.2. Are the epidemiology and animal bioassay study criteria/considerations scientifically justified and clearly described?

The criteria employed by the EPA in assessing the appropriateness of the available studies for use in the dose-response analysis are clearly stated for both the epidemiological and animal studies (Figure ES-1 and ES-2). The specific evaluation of each available study utilizes these criteria to make the argument for inclusion or exclusion of the study in the dose-response analysis. The five criteria constitute excellent guidelines in reaching these decisions, however two in particular require detailed information in order to evaluate the study's feasibility. These are discussed in detail below.

2.3. Has EPA applied the epidemiology and animal bioassay study criteria/considerations in a scientifically sound manner? If not, please identify and provide a rationale for alternative approaches.

The five criteria constitute excellent guidelines in reaching these decisions, however two in particular require detailed information in order to evaluate the study's feasibility. These two criteria are: 1) Confounding and other potential sources of bias are addressed, and 2) Statistical precision, power, and study follow-up are sufficient. From the description provided for some of the studies, it appears that these criteria were not consistently applied; specific examples are

included below. An additional criterion for potential inclusion of a study was: “The study is published in peer-reviewed literature *and includes an appropriate discussion of strengths and limitations*” (page 2-27, lines 19-20). It is not clear how the appropriateness of these discussions was determined, but this reviewer believes that different aspects of strength and weakness are likely to be identified by different reviewers at times. Thus allowing these papers with perhaps incomplete reviews of strengths and weaknesses to be evaluated further if they meet the other criteria should be permitted.

Re: Confounding and other potential sources of bias are addressed

The differences between males and females with regard to TCDD half-life are discussed, but the description of the number of males and females in each study population are often missing or very difficult to track down. Also, in the occupational cohort studies, the possibility of men and women performing different job tasks also increases the possibility that the men and women were exposed at different levels. However, when the job categories with assigned TCDD exposure levels are presented, there is often no discussion of the numbers by gender in the categories. For example, the Manz et al. study (1991) of the Hamburg cohort (1,583 men and 399 women) does not describe the TCDD categories by gender. In addition, the validity of the TCDD exposure levels assigned to the categories was examined “in a group of 48 workers who provided adipose tissue samples.” (Page 2-41, lines 18-19). How were these workers selected? How many were approached but refused to provide a sample? Assessment of selection bias in this and other similar circumstances is lacking in some of the studies. This is particularly notable in the lack of overall response rates reported for several of these studies. Inclusion of these factors in the study review would be very helpful.

RE: Statistical precision, power, and study follow-up are sufficient

This can be difficult to determine with the smaller sample size populations, but there are studies that can be very useful even given the small samples. For example, the relative risks calculated for increasing TCDD exposure and risk of breast cancer in the Seveso study were greatly increased in the 3rd and 4th highest exposure categories, but the RRs were not statistically significant (page 2-56, lines 1-8). However, as the EPA document states: “Although statistical significance was not achieved for either category, likely because of the small number of cases, the greater than three-fold risk evident in both categories is worth noting.” This needs to be kept in mind for additional evaluations in other studies as well.

Section 4. Reference Dose

I believe the selection of Mocarelli (2008) and Baccarelli (2008) for the RfD derivation was scientifically justified and clearly described. I also approve of the selection of adverse male reproductive effects and neonatal thyroid hormone levels as co-critical effects for the RfD, as the emergence of early windows of susceptibility as the most critical periods of exposure in the life cycle would be reflected in these outcomes. Moreover, alteration of neonatal thyroid hormone levels have been associated with several adverse child health outcomes, including neurologic development, and behavioral and auditory dysfunction.

Since “Barker’s Hypothesis” was introduced approximately twenty-five years ago, it is widely recognized that the periconceptional and perinatal periods of the life cycle likely represent the most highly susceptible windows for adverse effects due to in utero nutritional environment

[Barker 1993, Barker 1995]. Recent progress has expanded this hypothesis to include fetal exposures to environmental toxic chemicals and the impact on subsequent adverse adult outcomes, including cardiovascular disease, obesity, diabetes, cancers, Parkinson's Disease, developmental immunotoxicity, and Alzheimer's Disease [Selevan et al., 2000; Makris, 2008; Dietert 2009; Bateson and Schwartz, 2004; Landrigan et al., 2005; Patz et al., 2005]. There is tremendous interest in evaluating the relationship between these periconceptionalexposures, which include TCDD, and adverse child and subsequent adult health effects. Exposures at much lower levels than those observed in adults with no adverse health effects could be highly detrimental during pregnancy due to the susceptibility of the developing organism (Rubin and Soto 2009). Given the validity of both the TCDD exposure assessment as well as the outcome measures in the Mocarelli and Baccarelli studies, the choice of assessing male reproductive effects and neonatal thyroid hormone levels is appropriate and scientifically justified.

If feasible, and sample size permitting, the categorization of the males in the Mocarelli study to ages 1-10 and 10-17 years may be revisited in terms of the possibility of assessing the prenatal exposure period of exposure in deciding upon an RfD.

Given the limited understanding of the relationship between maternal and neonatal thyroid hormone levels, use of reported maternal levels as the exposure estimate in the Baccarelli study appears to be the best approach.

Uncertainty Factors: The EPA's discussion of the uncertainty factors considered in this evaluation is scientifically valid and comprehensive. Complex issues limiting exposure assessment, especially in the epidemiological studies, were well described and addressed. They acknowledged the importance of examining early windows of susceptibility while trying to account for differences in exposure during the interval until an adverse health effect is investigated. Other host factors that may influence these relationships, e.g., gender, BMI, smoking and other lifestyle factors, as well as concurrent exposures to DLCs, are all taken into consideration. The power of some existing studies to assess these issues is also restricted, particularly when examining early windows of susceptibility results in small numbers per age categories.

Sweeney Comments – Updated August 27, 2010

General Comments: I would like to extend my gratitude to the EPA researchers for their impressive efforts in responding to the NAS concerns. Of particular note is the active inclusion of the public in both the identification of additional sources of dioxin-related materials and in participation in the February workshop and in the current Dioxin Panel Review activities.

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis

2.1. Is this section responsive to the NAS concern about transparency and clarity in data-set selection for dose-response analysis?

I believe that the EPA has greatly improved their approach in addressing these concerns. The EPA's collaboration with Argonne National Laboratory and invitation to the public to engage in updating the literature search to identify all appropriate studies for evaluation, as well as the conduct of the Dioxin Workshop in February of 2009, were instrumental in enhancing the transparency and clarity regarding the process of selection of studies for the dose-response analysis. The development of clear criteria for study evaluation and inclusion (discussed below) were crucial in resolving the concerns raised by the NAS.

2.2. Are the epidemiology and animal bioassay study criteria/considerations scientifically justified and clearly described?

The criteria employed by the EPA in assessing the appropriateness of the available studies for use in the dose-response analysis are clearly stated for both the epidemiological and animal studies (Figure ES-1 and ES-2). The specific evaluation of each available study utilizes these criteria to make the argument for inclusion or exclusion of the study in the dose-response analysis. The five criteria constitute excellent guidelines in reaching these decisions, however two in particular require detailed information in order to evaluate the study's feasibility. These are discussed in detail below.

2.3. Has EPA applied the epidemiology and animal bioassay study criteria/considerations in a scientifically sound manner? If not, please identify and provide a rationale for alternative approaches.

The five criteria constitute excellent guidelines in reaching these decisions, however two in particular require detailed information in order to evaluate the study's feasibility. These two criteria are: 1) Confounding and other potential sources of bias are addressed, and 2) Statistical precision, power, and study follow-up are sufficient. From the description provided for some of the studies, it appears that these criteria were not consistently applied; specific examples are included below. An additional criterion for potential inclusion of a study was: "The study is published in peer-reviewed literature *and includes an appropriate discussion of strengths and limitations*" (page 2-27, lines 19-20). It is not clear how the appropriateness of these discussions was determined, but this reviewer believes that different aspects of strength and weakness are likely to be identified by different reviewers at times. Thus allowing these papers with perhaps incomplete reviews of strengths and weaknesses to be evaluated further if they meet the other criteria should be permitted.

Re: Confounding and other potential sources of bias are addressed

The differences between males and females with regard to TCDD half-life are discussed, but the description of the number of males and females in each study population are often missing or very difficult to track down. Also, in the occupational cohort studies, the possibility of men and women performing different job tasks also increases the possibility that the men and women

were exposed at different levels. However, when the job categories with assigned TCDD exposure levels are presented, there is often no discussion of the numbers by gender in the categories. For example, the Manz et al. study (1991) of the Hamburg cohort (1,583 men and 399 women) does not describe the TCDD categories by gender. In addition, the validity of the TCDD exposure levels assigned to the categories was examined “in a group of 48 workers who provided adipose tissue samples.” (Page 2-41, lines 18-19). How were these workers selected? How many were approached but refused to provide a sample? Assessment of selection bias in this and other similar circumstances is lacking in some of the studies. This is particularly notable in the lack of overall response rates reported for several of these studies. Inclusion of these factors in the study review would be very helpful.

RE: Statistical precision, power, and study follow-up are sufficient

This can be difficult to determine with the smaller sample size populations, but there are studies that can be very useful even given the small samples. For example, the relative risks calculated for increasing TCDD exposure and risk of breast cancer in the Seveso study were greatly increased in the 3rd and 4th highest exposure categories, but the RRs were not statistically significant (page 2-56, lines 1-8). However, as the EPA document states: “Although statistical significance was not achieved for either category, likely because of the small number of cases, the greater than three-fold risk evident in both categories is worth noting.” This needs to be kept in mind for additional evaluations in other studies as well.

Section 4. Reference Dose

I believe the selection of Mocarelli (2008) and Baccarelli (2008) for the RfD derivation was scientifically justified and clearly described. I also approve of the selection of adverse male reproductive effects and neonatal thyroid hormone levels as co-critical effects for the RfD, as the emergence of early windows of susceptibility as the most critical periods of exposure in the life cycle would be reflected in these outcomes. Moreover, alteration of neonatal thyroid hormone levels have been associated with several adverse child health outcomes, including neurologic development, and behavioral and auditory dysfunction.

Since “Barker’s Hypothesis” was introduced approximately twenty-five years ago, it is widely recognized that the periconceptional and perinatal periods of the life cycle likely represent the most highly susceptible windows for adverse effects due to in utero nutritional environment [Barker 1993, Barker 1995]. Recent progress has expanded this hypothesis to include fetal exposures to environmental toxic chemicals and the impact on subsequent adverse adult outcomes, including cardiovascular disease, obesity, diabetes, cancers, Parkinson’s Disease, developmental immunotoxicity, and Alzheimer’s Disease [Selevan et al., 2000; Makris, 2008; Dietert 2009; Bateson and Schwartz, 2004; Landrigan et al., 2005; Patz et al., 2005]. There is tremendous interest in evaluating the relationship between these periconceptional exposures, which include TCDD, and adverse child and subsequent adult health effects. Exposures at much lower levels than those observed in adults with no adverse health effects could be highly detrimental during pregnancy due to the susceptibility of the developing organism (Rubin and Soto 2009). Given the validity of both the TCDD exposure assessment as well as the outcome measures in the Mocarelli and Baccarelli studies, the choice of assessing male reproductive effects and neonatal thyroid hormone levels is appropriate and scientifically justified.

Regarding the derivation of the reference dose, however, there is a grave concern regarding the utilization of a sufficiently high dose at critical stages of susceptibility *alone*, e.g., periconceptional and gestational exposures regardless of subsequent exposures, versus the need to consider these later exposures. This also reinforces the need to consider that adverse effects due to identical doses during these highly susceptible windows of the life cycle and doses at older ages could be vastly different.

If feasible, and sample size permitting, the categorization of the males in the Mocarelli study to ages 1-10 and 10-17 years may be revisited in terms of the possibility of assessing the prenatal exposure period of exposure in deciding upon an RfD.

Given the limited understanding of the relationship between maternal and neonatal thyroid hormone levels, use of reported maternal levels as the exposure estimate in the Baccarelli study appears to be the best approach.

Uncertainty Factors: The EPA's discussion of the uncertainty factors considered in this evaluation is scientifically valid and comprehensive. Complex issues limiting exposure assessment, especially in the epidemiological studies, were well described and addressed. They acknowledged the importance of examining early windows of susceptibility while trying to account for differences in exposure during the interval until an adverse health effect is investigated. Other host factors that may influence these relationships, e.g., gender, BMI, smoking and other lifestyle factors, as well as concurrent exposures to DLCs, are all taken into consideration.

The power of some existing studies to assess these issues is also restricted, particularly when examining early windows of susceptibility results in small numbers per age categories. Another major limitation is the restriction of studies to those evaluating solely TCDD exposure. Numerous studies that involve dioxin-like exposures could be included in this effort by taking these co-exposures into account in the analysis.

Finally, regarding uncertainty factors, a strong case can be made to examine exposures occurring during the periconceptional and gestational time intervals. The uncertainties inherent in trying to reconstruct earlier exposures if assessment begins in late childhood/adulthood are daunting. And again, similar doses may be highly detrimental during these time periods but have no negative impacts on older individuals. In this respect, the emphasis would be placed on the most susceptible subpopulations, which ensures that the population as a whole would be adequately protected.

Dr. Mary Walker

Preliminary Comments – July 7, 2010

Charge Question

1.1. “clear and logical?”

- The NRC review listed eight conclusions and recommendations, while three key areas were noted as requiring substantial improvement. It is not clear why EPA chose to address only the three key areas of concern, and if, and when the other concerns will be addressed.
- It is not clear if, and when, EPA will address the scientific input that they received from the dioxin workshop held in February 2009.
- My concerns, comments, and confusion do not necessarily “fit” into the assigned charge questions. I have organized my preliminary responses by section and by question when possible.

1.2 “other critical studies?”

- No specific recommendations at this time

Section 2 “Transparency and clarity in data-set selection for dose-response modeling?”

- It is not clear how “standard toxicological practices” is defined. Further, it is not clear if there were data-sets specifically excluded based on this criterion?
- Data set selection for noncancer studies appears to have gone through a two cycle screening process that is not clearly outlined. In order to make the decision to exclude studies with doses >30 ng/kg/d, the EPA first had to identify studies for noncancer endpoints that had low LOAELs (i.e. ~ 1.0 ng/kg/d). Is this correct?
- While the inclusion/exclusion criteria are clearly delineated for data-sets selected to be reviewed and considered for dose-response modeling, it is significantly less transparent why certain studies were then not used to calculate an RfD. The manner in which the document is organized makes it extremely difficult to track a given study from its initial selection in Chapter 2 through to RfD calculation in Chapter 4. For example, some selected data sets in Chapter 2 do not appear on Table 4-3 in Chapter 4 and there are multiple possible reasons why these studies were not carried forward: (a) endpoints were not considered toxicological relevant, (b) other DLCs could be making a significant contribution to the toxicological endpoint, (c) results are not reproducible by another study, (d) route of exposure was not relevant (i.e. ip rather than oral), (e)?? At least two studies, Hochstein et al. 2001 and Sugita-Konishi et al. 2003, represent selected data-sets in Chapter 2, but are excluded from Table 4-3. The rationale for their exclusion is not clearly explained (Hochstein) or is inappropriate (Sugita-Konishi). For Sugita-Konishi et

al. 2003, toxicological adverse endpoints, including reduced spleen weight and decreased clearance rate of a bacterial infection were significantly altered by TCDD exposure. Excluding the TNF α and IFN γ responses may be acceptable. There needs to be a more systematic manner for tracking data-sets from initial selection through additional exclusion criteria to dose-response modeling to RfD calculation.

- This reviewer identified discrepancies between original published literature and the summary descriptions in Chapter 2. These discrepancies included errors in stating the appropriate adverse outcome and in dose calculation, which in some cases were carried forward to the dose-response modeling. Given the short timeframe available for review, it was not possible to conduct a systematic comparison for all selected data sets with the original literature; however, the lack of accuracy strongly suggests that a thorough QA/QC needs to be conducted.

Section 4 “Reference Dose”

- In Figure 4-2, the definition of “minimum LOAEL” is not clear. This is presumably the lowest LOAEL reported for all endpoints studied in noncancer animal bioassays. If so, what value and endpoint were used?
- Question 4.1. The results from the animal bioassays to the RfD calculation are strictly qualitative and not used in any manner to establish the RfD. This assumes that the endpoints measured in the two co-critical human epidemiology studies are the most sensitive endpoints of concern for human exposure and are “the critical effect”. This assumption is questionable since not all possible endpoints have been assessed in humans, which is the reason why animal studies are conducted.
- Question 4.4. Assessing the degree to which an endpoint is adverse is fundamental to reducing uncertainty associated with risk assessment. Changes in biochemical endpoints could be as important in a toxicological response as changes in hormone levels, for example. As long as the biochemical response has been linked to an adverse (i.e. pathological) event in the literature in the tissue of measurement, then it should be considered adverse even if not linked to an adverse response in the current study. This requires a thorough review of the literature relative to the biochemical change being measured, because a study identifying a biochemical change could be conducted years before that change has been linked to some adverse event. While EPA’s rationale for excluding specific studies measuring certain biochemical endpoints following TCDD exposure appears reasonable, the original papers would need to be reviewed to evaluate the degree and duration of response and its potential to be associated with an adverse response. [A discussion of excluding Selzac et al. 2000 needs to be added to appendix G.]
- Question 4.8. Table 4-6. The strengths and limitations associated with the animal bioassays require clarification. Before providing a narrative of strengths and weaknesses of individual studies, the important criteria against which they are judged need to be clearly established. What is considered a large sample size versus small? How is this evaluation judged based on the endpoint being measured? An endpoint that exhibits low biological variability does not require a large sample size to

establish a statistically significant difference. If qualitative analysis of strength includes comparison to the endpoint being measured in humans, this comparison should be stated for all studies. For example, if the NTP 2006 study is given a lot of weight or is considered a gold standard, it would be important to know whether the endpoints measured in this study were ever reported in humans and at what exposure levels.

- Question 4.8. Simply because the rodent PBPK model is a poor choice for dose-response modeling for the mouse is not an adequate justification for dismissing the results of the mouse studies.
- Question 4.8. It is not clear why all studies for which a candidate RfD was calculated were not discussed in Table 4-6.

Walker Comments Updated – August 20, 2010

Responses to Charge questions on “EPA’s Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments”

Section 2. Transparency and Clarity in the Selection of Key Data Sets for Dose-Response Analysis

2.1 Is this Section responsive to the NAS concern about transparency and clarity in data-set selection for dose-response analysis?

Answer: Only partially. The inclusion/exclusion criteria for data-set selection are clearly delineated. However, *after a data-set was selected* it is not transparent what additional criteria were used to include or exclude it for dose-response modeling and then to include or exclude it for RfD calculation. As a result it is very difficult to track a given study from its initial selection in Chapter 2 through to RfD calculation in Chapter 4. For example, some selected data sets in Chapter 2 do not appear on Table 4-3 in Chapter 4 and there are multiple possible reasons why these studies were not carried forward: (a) endpoints were not considered toxicological relevant, (b) other DLCs could be making a significant contribution to the toxicological endpoint, (c) results are not reproducible by another study, (d) route of exposure was not relevant (i.e. ip rather than oral), (e)?? These criteria need to be equally transparent and clear as those for data-set selection and currently they are not. At least two studies, Hochstein et al. 2001 and Sugita-Konishi et al. 2003, represent selected data-sets in Chapter 2, but are excluded from Table 4-3. The rationale for their exclusion is not clearly explained (Hochstein) or is inappropriate (Sugita-Konishi). For Sugita-Konishi et al. 2003, toxicological adverse endpoints, including reduced spleen weight and decreased clearance rate of a bacterial infection were significantly altered by TCDD exposure.

In addition, one of the data-set selection criteria needs to be defined in more specific terms; “standard toxicological practices”. It is not clear how this criterion was defined and if there were data-sets specifically excluded based on it? Perhaps these standard toxicological practices have been delineated in previous EPA documents and if so, then those documents should be cited.

Thus, although the transparency and clarity of data-set *selection* is very good, the subsequent decision making process and criteria for inclusion or exclusion for dose-response modeling and then RfD calculation is not as clear.

2.2 Are the epidemiology and animal bioassay study criteria/considerations scientifically justified and clearly described?

Answer: One criterion that is applied after data-set selection, but before dose-response modeling is the decision to include only those studies in which exposure was “primarily” to TCDD and exclude those studies in which the contribution of DLCs was significant or cannot be adequately assessed. Although the scientific rationale for this criterion is probably reasonable, the criterion itself is overly vague, must be applied equally across all selected data-sets, and the rationale for the criterion must be discussed in greater detail. The lack of application of this criterion equally across all data-sets is particularly concerning. For example, an animal study of monkeys (Rier et al. 2001, 199843) is excluded because the contribution of serum TCDD to the total serum TEQ was 30%, while one of the co-critical epidemiology studies of humans (Baccarelli et al 2008) the contribution of serum TCDD to the total serum TEQ was 46% in the exposed group; representing a difference of 16%. It is not clear that this difference is biologically meaningful and that either study results in exposure “primarily” to TCDD. Thus, “primarily” exposed to TCDD needs to be more clearly defined and applied equally across animal bioassay and human epidemiology data-sets.

It is also worth noting that EPA has chosen to exclude most epidemiology studies of Vietnam veterans in which it has been clearly shown that their exposure was primarily to TCDD and that serum levels of DLCs are equivalent between exposure and comparison groups (Pavuk *et al.*, 2007). If this is an important criterion, then the Vietnam veterans studies likely can be useful for this risk assessment.

2.3 Has EPA applied the epidemiology and animal bioassay study criteria/considerations in a scientifically sound manner? If not, please identify and provide rationale for alternative approaches.

Answer: Not in all cases. The decision to exclude epidemiology studies of Vietnam veterans (specifically Michalek and Pavuk 2008, 199573) is not scientifically justified. The rationale for exclusion (page 2-124, lines 18-19) “the possible confounding from the inability to control for 2,4-D and other agents used in Agent Orange precludes a quantitative dose-response analysis” does not seem reasonable and appears to apply a double standard compared to use of data from the NIOSH cohort. It is reasonable to assume that individuals from the NIOSH cohort will have been exposed to similar agents, including 2,4-D and 2,4,5-T, and likely similar levels as Vietnam veterans. The rationale for exclusion is inadequate and discards some very important and well controlled studies. Further, there are at least a couple of epidemiology studies of Vietnam veterans that were not considered for non-cancer endpoints in Chapter 2 and the rationale for their exclusion is not provided. These include (Michalek *et al.*, 1999) and (Henriksen *et al.*, 1997).

In addition, the decision to exclude the Kang et al. 2006, 199133 also is not scientifically justified. The rationale for exclusion (page 2-130, lines 4-5) “the lack of demonstrated dose-response relationships with cancer or other outcomes precluded the use of these data for characterizing the dose-response from TCDD.” Again, this appears to apply a double standard

compared to use of data from other epidemiology studies, specifically one of co-critical studies Mocarelli et al. 2008. The latter study compares an exposed group to a comparison group similar to the Kang et al. 2006 study. It is unclear why one study is included, while the other is excluded. While data limitations may exist, these need to be more clearly explained.

Lastly, discrepancies occur between original published literature and the summary descriptions in Chapter 2. These discrepancies include errors in stating the appropriate adverse outcome and in dose calculation, which in some cases were carried forward to the dose-response modeling. The EPA is strongly encouraged to conduct a thorough review and quality control of the entire document in order to ensure complete accuracy and use of the reported data.

Section 4. Reference Dose

4.1 Is the rationale for the selection of Mocarelli et al. (2008) and Baccarelli et al. (2008) as co-critical studies for derivation of the RfD scientifically justified and clearly described? Please identify and provide the rationale for any other endpoints that should be selected as the critical effect.

Answer: While individuals (including myself) attending the Feb 2009 workshop sent the message that use of human data would be preferred over animal data in derivation of an RfD, in my opinion this meant when high quality human studies were available. The two human studies identified as co-critical are valuable, but have some significant limitations and are not as defensible in my opinion as using the best rodent study available for setting the RfD.

Furthermore, the results from the animal bioassays to the RfD calculation are strictly qualitative and not used in any manner to establish the RfD. This assumes that the endpoints measured in the two co-critical human epidemiology studies are the most sensitive endpoints of concern for human exposure and are “the critical effect”. This assumption is questionable since not all possible endpoints have been assessed in humans. Again, I would recommend using the best rodent study available for setting the RfD and then support this number by comparing it to human studies for which an RfD can be calculated.

One of the weaknesses of the document is the inadequate discussion of how animal studies and observations in humans are similar or different. The two co-critical human epidemiology studies as they stand alone are not very convincing and could be strengthened (or weakened) by a clear weight-of-evidence discussion that includes experimental animal studies.

Lastly, an adequate and convincing justification for why the endpoints in these two co-critical studies should be considered adverse and disease-associated is lacking. This is an essential requirement if these two studies are going to serve as the basis for an RfD determination.

4.2 a.i Please comment on EPA’s approach for identifying the exposure window and calculating the average exposure for the Mocarelli et al. 2008 study.

Answer: The exposure window is broad and there are no data available to scientifically justify how the exposure should be calculated. One approach would over estimate, while the other would under estimate. Taking an average of the two is simplistic and not scientifically defensible. Thus, this decision is a risk management decision, not a scientific one.

4.2.a.ii. Please comment on EPA's designation of a 20% decrease in sperm count as a LOAEL.

Answer: It is not clear whether the 20% decrease in sperm count is a reasonable LOAEL and I did not find that adequate justification was provided. It may be more appropriate to consider the absolute values to convincingly justify whether this decrease in sperm count would be biologically meaningful, rather than using the percentage decrease. Since the recommendations of the WHO were being considered for the Bacarelli et al. 2008 study, I also considered those recommendations when evaluating the Mocarelli et al. 2008 study. In a recent publication WHO determined reference values for human semen characteristics from fertile men whose partners had conceived in ≤ 12 months (Cooper *et al.*, 2010). The range of sperm count/ejaculate within the 25-75th centiles was $142-422 \times 10^6$ sperm (n=1859) and the reference limit, defined as the sperm count at the 5th centile, was 39×10^6 sperm (95% confidence interval, $33-46 \times 10^6$ sperm). The mean sperm count reported in the Mocarelli et al. 2008 study was 149.8×10^6 sperm in the exposed group (n=71) and 186.1×10^6 sperm in the comparison group (n=82). Thus, although the exposed group exhibits a 20% reduction in sperm count, comparing the absolute values to the WHO reference values suggests that both groups are well within the normal range for fertile men. Thus, it is not clear nor justified that the 20% decrease in sperm count in these exposed individuals should be considered adverse.

4.2.b.i. Please comment on EPA's decision to use the reported maternal levels and the appropriateness of this exposure estimate

Answer: Use of the maternal levels rather than soil exposure is very appropriate.

4.2.b.ii Please comment on EPA's designation of 5 μ -units TSH per ml blood as a LOAEL.

Answer: It is important to address the manner in which the TSH testing was conducted. Two variables that can significantly influence the level of measured TSH include age of the neonate and method used for analysis (Gruneiro-Papendieck *et al.*, 2004). First, while it is true that newborns have high TSH levels immediately after birth, it is well documented that these levels stabilize after 48 hrs and thus the WHO guidelines specifically state that screening should be conducted after 48 hr and up to 3 weeks post parturition. In Bacarelli et al. 2008, the TSH screening was conducted 3 days after birth, which is in complete accordance with WHO guidelines. Second, two standard assay methods for TSH analysis are immunofluorescence and immunoradiometric. The immunofluorescence assay has been found to be more consistent and accurate for samples with concentrations $< 5 \mu\text{U/ml}$. The Bacarelli et al. 2008 paper used the immunofluorescence assay. Lastly, current WHO guidelines consider that $< 3\%$ of newborns should have neonatal TSH levels of $< 5 \mu\text{U/ml}$ blood to be considered iodine sufficiency. Thus, there is no requirement that the population averages $< 5 \mu\text{U/ml}$ as suggested by an external commenter. The methods used to analyze the neonatal TSH levels are within normal standard practices and are appropriate.

Nonetheless, I think there are some limitations to using this endpoint. First, the WHO guidelines (1994) for neonatal TSH levels are quite specific and are interpreted incorrectly by EPA (page 2-118, lines 7-8). WHO states that the *frequency* of TSH levels $> 5 \mu\text{U/ml}$ must exceed 3% to be considered indicative of iodine deficiency, *not simply that a TSH level $> 5 \mu\text{U/ml}$ in an individual neonate is considered "bad"*. By these criteria, the frequency of TSH levels $> 5 \mu\text{U/ml}$ is elevated in both Zone A and B, and right on the edge for the reference group. The problem is that there is no evidence provided in the Bacarelli et al. 2008 paper that a mild iodine

deficiency did not exist, which potentially confounds the ability to detect an association with TCDD exposure. [Although zone A which has the highest soil TCDD concentration still demonstrates a frequency of TSH levels $> 5\mu\text{U/ml}$ of 16% (9/56), EPA does not use these data since exposure misclassification could occur and instead relies on maternal levels for an exposure assessment.] When individual maternal exposure is considered (Fig 2A), the sample size drops to $n=51$. If a mild iodine deficiency is present in this population, then 1.5 neonates would be predicted to have a TSH levels $> 5\mu\text{U/ml}$. It is notable that 2 neonates had TSH levels $> 5\mu\text{U/ml}$ and their mothers had very low serum TCDD levels (3-4 ppt). In contrast, 5 neonates had TSH levels $> 5\mu\text{U/ml}$ and all their mothers had serum TCDD levels of ≥ 50 ppt. Since there was no significant correlation between maternal serum TCDD concentration < 50 ppt and neonatal TSH levels, it would be most appropriate to set the LOAEL at 50 ppt rather than at the geometric mean of TCDD for all individuals with TSH $> 5\mu\text{U/ml}$. Further, 50 ppt should represent the LOAEL, not the NOAEL, since 5/6 neonates (83%) born to mothers with serum TCDD concentration ≥ 50 ppt exhibited an effect. Nonetheless, this is an extremely small sample size ($n=5$) to use as a co-critical study and serve as the basis to calculate an RfD.

The next concern is the degree to which this endpoint should be considered adverse. As noted above, a TSH level $>5\mu\text{U/ml}$ in ONE individual neonate is not considered bad, but rather that a TSH level $>5\mu\text{U/ml}$ in a population of neonates is bad ONLY if $>3\%$ have an elevation. This frequency of elevation is associated with iodine deficiency. It has not been proposed that TCDD exposure induces iodine deficiency. Thus, the EPA needs to clearly justify what disease state or adverse outcome is likely to be associated with this increase in TSH.

4.3 Please comment on the rationale for the selection of the uncertainty factors for the RfD

Answer: The rationale for the selection of uncertainty factors is appropriate. However, EPA should consider including an uncertainty factor for data quality.

4.4 Please comment on whether the decision to exclude “biochemical” endpoints is scientifically justified and clearly described.

Answer: Assessing the degree to which an endpoint is adverse is fundamental to reducing uncertainty associated with risk assessment. Changes in biochemical endpoints could be as important in a toxicological response as changes in hormone levels, for example. The more important issue is whether the biochemical response has been linked to an adverse, disease-related event in the tissue in question. If this link has been established in the literature and is scientifically defensible, then it should be considered adverse even if it was not linked to an adverse response in the study being reviewed. This requires a thorough review of the literature relative to biochemical changes being measured, because a study identifying a biochemical change could be conducted years before that change has been linked to some adverse event. While EPA’s rationale for excluding specific studies measuring certain biochemical endpoints following TCDD exposure appears reasonable, the original papers need to be reviewed to evaluate the degree and duration of response and its potential to be associated with an adverse response. [A discussion of excluding Selzac et al. 2000 needs to be added to appendix G.] For example, while it has recently been demonstrated that CYP1A1 is required to mediate TCDD-induced vascular dysfunction and hypertension (Kopf *et al.*, 2010), a study measuring the TCDD dose-related induction of hepatic CYP1A1 may not be biologically relevant to the vascular toxicity. Nonetheless, it would be useful for general comparison if RfD values were calculated

based on these excluded biochemical endpoints to illustrate the sensitivity of changes that occur preceding the development of documented adverse responses.

4.5 Please comment on EPA's approach for averaging exposures.

Answer: Given the diversity of exposure scenarios used in animal bioassays, the EPA has chosen a very reasonable approach for determining exposure.

4.8 Please comment as to whether EPA's quantitative discussion of uncertainty in the RfD is justified and clearly described.

Answer: Table 4-6. The strengths and limitations associated with the animal bioassays require clarification. Before providing a narrative of strengths and weaknesses of individual studies, the important criteria against which they are judged need to be clearly established. What is considered a large sample size versus small? How is this evaluation judged based on the endpoint being measured? An endpoint that exhibits low biological variability does not require a large sample size to establish a statistically significant difference. If qualitative analysis of strength includes comparison to the endpoint being measured in humans, this comparison should be stated for all studies. For example, if the NTP 2006 study is given a lot of weight or is considered a gold standard, it would be important to know whether the endpoints measured in this study were ever reported in humans and at what exposure levels. It is not clear why all studies for which a candidate RfD was calculated were not discussed in this Table.

The definition of "minimum LOAEL" is not clear. This is presumably the lowest LOAEL reported for all endpoints studied in noncancer animal bioassays. If so, what value and endpoint were used?

While it was recommended that the application of the rat PBPK model to the mouse should be peer-reviewed, overall the EPA SAB was very positive about the use of the model in the mouse. However, in section 4, the results of the mouse studies are largely dismissed because EPA concludes that the rodent PBPK model is a poor choice for dose-response modeling for the mouse. This is clearly not an adequate justification for dismissing the results of the mouse studies.

Reference cited

Cooper, T. G., Noonan, E., von Eckardstein, S., Auger, J., Baker, H. W., Behre, H. M., Haugen, T. B., Kruger, T., Wang, C., Mbizvo, M. T. and Vogelsong, K. M. (2010). World Health Organization reference values for human semen characteristics. *Hum Reprod Update* **16**, 231-45.

Gruneiro-Papendieck, L., Chiesa, A., Mendez, V., Bengolea, S. and Prieto, L. (2004). Neonatal TSH levels as an index of iodine sufficiency: differences related to time of screening sampling and methodology. *Horm Res* **62**, 272-6.

Henriksen, G. L., Ketchum, N. S., Michalek, J. E. and Swaby, J. A. (1997). Serum dioxin and diabetes mellitus in veterans of Operation Ranch Hand. *Epidemiology* **8**, 252-258.

Kopf, P. G., Scott, J. A., Agbor, L. N., Boberg, J. R., Elased, K. M., Huwe, J. K. and Walker, M. K. (2010). Cytochrome P4501A1 is Required for Vascular Dysfunction and Hypertension Induced by 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicol Sci*.

Michalek, J. E., Akhtar, F. Z. and Kiel, J. L. (1999). Serum dioxin, insulin, fasting glucose, and sex hormone-binding globulin in veterans of Operation Ranch Hand. *J Clin Endocrinol Metab* **84**, 1540-3.

Pavuk, M., Patterson, D. G., Jr., Turner, W. E., Needham, L. L. and Ketchum, N. S. (2007). Polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), and dioxin-like polychlorinated biphenyls (PCBs) in the serum of US Air Force veterans in 2002. *Chemosphere* **68**, 62-8.