

Summary of Discussion in Response to Charge Questions

(As Discussed at the July 13-15, 2010 Meeting of the Dioxin Review Panel)

Section 2: Transparency and clarity in the selection of key data sets for dose-response analyses (Summary prepared by Dr. Lawrence)

On-balance this section of the Response has many aspects that are strong. The report goes to great lengths to document and justify the selection of key datasets used for the analysis. However, there was a general belief among SAB members that the Report would benefit from greater clarity regarding the data sets *excluded* at various steps in the selection process.

The following points for further discussion/suggestions for improvement were made (and are presented here in no particular order):

Overall:

- Again, we reiterate that there were many positive comments regarding the thoroughness and clarity of this section.
- However, this Section was generally viewed as overly verbose. Careful editing to revise and consolidate this section (and the document as a whole) are recommended.
- This Section could be structured such that it is easier to follow a study from one section of the document to another; in other words, improve overall document integration, using Section 2 as the foundation for this integration.
- There was extensive discussion surrounding the idea that key studies were kept in or excluded based on the following criteria: low dose threshold met, TCDD as the primary toxicant, and detailed dose-response measures including low doses. This information needs to be explicitly integrated in the appropriate tables and perhaps also briefly stated at the beginning of each Section of the document to improve clarity. This is particularly relevant for issues surrounding weight of evidence.
- EPA staff were asked to provide a more expanded rationale for their definitions of adverse health effects.
- The limitations of the narrow selection of just a few studies are a concern that could be addressed in appropriate sections. For example, during this meeting the EPA staff were asked to comment on systematic information gathering and culling. The gist of the question was to better understand why studies were excluded. In their response, EPA staff explained that they “weeded down the universe of studies” into a manageable amount in a tiered assessment for inclusion/exclusion. In the first pass, they did not always document why an individual study was excluded; however, for studies that made this first cut, EPA staff then systematically documented the basis for exclusion. This overall approach seemed to make sense to the committee. However, SAB members conveyed that it is difficult to glean this from the current document. A recommendation was made that this information be added to Table 2.7 (by EPA) to make this process much more evident to an outside EPA reader. To be clear, the goal is to get EPA to better

explain how and why studies were excluded, and the implications of such inclusions/exclusions for RfD derivations in Sections 4 and 5.

Other discussion and suggestions regarding inclusion/exclusion criteria were as follows:

- The rationale for distinct criteria for epidemiological and animal studies could be made stronger, and data set selection for non-cancer and cancer endpoints has room for further clarification and justification.
- There does not appear to be consensus regarding the scientific justification for some of the inclusion and exclusion criteria; with majority of opinions leaning toward a sense that excellent studies were excluded for reasons that are not well justified.
 - Sense that perhaps EPA may have been too stringent in exclusion of some excellent studies. Suggestions were made by several SAB members that adding information to the appendices and/or tables would provide readers with clarification regarding the exclusion of particular studies.
 - Awareness of other studies was expressed by several, with a mixed sense of whether including them would (or would not) have a significant impact on the dose-response assessment

Other discussion and suggestions

- The requirement that TCDD purity be explicitly noted in a publication needs justification
- Some phrasing could use more explicit definition. For example, ‘*study design is consistent with standard toxicological practices*’ is vague and subject to considerable bias. Another example is the phrase “consistent with EPA criteria or policy.” The opinion was expressed that this statement is neither adequate nor fully responsive to NAS criticisms and suggestions for improvement. The scientific basis for a particular justification or decision needs to be stated.
- The choice to exclude studies of DLCs was not fully supported by all SAB members, and discussion of this decision in the report could be given more weight.
- The weight of evidence given to null studies (compared to studies that show a correlation) could be explained in more detail. Specifically the power to detect significant differences for rare events should be considered.

Section 3. The Use of Toxicokinetics in the Dose-Response Modeling for Cancer and Noncancer Endpoints. (Summary provided by Dr. Clewell)

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 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 than body burden, because it is more closely related to the biologically relevant dose metric: the free concentration of dioxin in the target tissues (liver, fetus, etc.). Blood concentrations are routinely used to estimate biologically effective exposures for pharmaceuticals.

The rationale for the use of whole blood concentration rather than lipid adjusted serum concentration (LASC) should not be based on the Emond model structure. It would be trivial to change the model so that LASC could be predicted. Indeed, the model is apparently used to estimate LASCs in the RfD calculations (e.g., p. xli, line 21). 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. A metric that considers blood lipid content 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 this issue is addressed in the dose metric calculations. 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.

The Emond model provides the best available basis for the dose metric calculations in the assessment. It is the product of a high-caliber, multi-year research effort at EPA/NHEERL led by Linda Birnbaum and Mike Devito, and represents a significant effort in terms of data collection. However, additional discussion of other published models and quantitative evaluation of the impact of model selection on dose metric predictions should also be provided.

This discussion should also address how the model is intended be used in the assessment, which would then dictate why a particular model was selected. That is, for the intended purposes, was the Emond model more robust and/or simpler than other models, and did it contain sufficient details for biological determinants deemed important by the Agency.

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, 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.

3.2. 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).

An appropriate approach was used to develop the mouse model on the basis of the published rat model and the available mouse kinetic data. Nevertheless, an external peer review of the mouse model should be performed, since this model has not been published in the peer-reviewed literature, which is typically a requirement for models to be used by the Agency.

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

The mouse model performs reasonably well, apart from under-prediction of urinary excretion data. The model appears to be adequate for use in estimating dose metrics for the assessment, but with greater uncertainty than the rat and human models. The EPA's suggestion in the RfD chapter that the clustering of mouse PODs at the lowest doses is due to model failure is inappropriate.

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 qualitative 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., Seveso). The modeling of the Cheng, Moccarelli, and Bacarelli studies needs to be described in more detail and the impact of model parameter uncertainty and exposure uncertainty in these studies should be evaluated quantitatively.

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

The EPA document only presents the sensitivity analysis published by Emond et al. 2006, which is not entirely adequate for the purposes of this assessment. It leaves out the Hill coefficient, which is one of the most important parameters in the model for low-dose extrapolation (Evans and Andersen, 2000). 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.

We agree with the average daily dose calculation approaches described in the EPA document, although the predictions of the model in the perinatal period need to also be evaluated for the possibility that the change in exposure associated with birth might lead to transient changes in peak blood concentration.

Section 4. Reference Dose (Summary to be provided by Dr. Faustman)

Section 5. Cancer Assessment (Summary provided by Dr. Håkansson)

In general, panel members were impressed by the extensive work presented by the Agency in their response to the NAS comments on cancer assessment. Comments below are supporting the Agency in further developing section 5 and to transfer some of its contents to other sections of the draft.

5.1. 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?

Response 5.1: Panel members agreed on the classification that “TCDD is carcinogenic to humans”. There is need of more information from the Agency on the power of studies used and the difficulties involved when assessing rare tumors. Thoroughly addressing these aspects will make the weight of evidence characterization in this section more clear and transparent. Other key issues which were discussed among panel members were the need of a more thorough and up-to-date mode-of-action discussion (see 5.2), the need to also consider studies based on the TEQ-dose metric (few human studies exist with a pure TCDD exposure; see also 5.3 and 5.7), and whether dioxin (TCDD/TEQ) is carcinogenic to humans at present background exposure levels. Panel members also pointed out that in the weight-of-evidence characterization it is important to build on all the available data to support the decision, even if in the end only one or two (human) studies will be selected for setting an RfD or similar. The argumentation needs to build on “all” available data and it needs to be clear how different types of data (in vitro, in vivo, human) support each other; or not.

5.2 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.

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.

Response to 5.2a,b: Panel members pointed out the need of a general dioxin mode-of-action paragraph, which covers both cancer and non-cancer events. Such a paragraph would fit better into an earlier section of the document e.g., section 4. Panel members pointed out that much is known about dioxin/TCDD toxicity and mode-of-action. Nevertheless, until now, the exact mechanism-of-action has not been fully delineated for any distinct TCDD-toxicity end-point. Panel members strongly supported that the Agency provides an up-to-date dioxin mode-of-action

section in its response to NAS comments, which takes into account recent developments in building formalized concepts to distinguish between mode-of-action(s) and mechanism(s)-of-action. It is important to make clear why (in the risk assessment steps for dioxin) there is a need to know even more about the mechanism-of-action for different toxicity end-points (e.g. very small or even non-existing margin of exposure; all individuals exposed from the earliest life-stages and throughout life). Life-stage sensitivity issues also need to come across clearly in the mode-of-action section.

Panel members appreciated the attempts by the Agency to further develop cancer mode-of-action concepts based on available dioxin liver, lung, and thyroid toxicity data. Such innovative and explorative work is clearly fundamental to the continued need of further developing risk assessment sciences and to make more detailed and integrated use of already existing and published data. As presented now, there is need of more text to explain data selections, methods used and conclusions drawn both for the general text and for the associated figures/tables.

5.3 Data selection. 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?

Response 5.3: Panel members proposed that the data selection approach presented in section 2 should be further developed to also cover exactly how section 5 studies were selected. It should be made very clear and visible already in table format, which studies were carried forward or not. Panel members discussed the need to include studies with dioxin-like compounds (DLC) in the evaluation and the possibility to include TEQs in table 2-1 (see also 5.7).

5.4 Animal bioassay data. 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.

Response 5.4a: .

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.

Response 5.4b: Panel members noted the consistency with Agency guidelines and had no further comments.

5.5 Cheng et al. 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.

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.

Response 5.5a: Panel members agreed that Cheng et al is the right study, and the selection of this study is well described.

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.

Response 5.5b: Panel members agreed that it is appropriate to use all-cancer mortality in this case, although generally this is not a good idea. Using all-cancer mortality makes it more difficult to get an association.

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.

Response 5.5c: Panel members agreed that the use of the Emond study is scientifically justified and clearly described.

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.

Response 5.5d: Panel members agreed that the Agency has chosen the appropriate extrapolation model and that using the oral slope factor to arrive at the POD was correctly done.

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.

Response 5.5e:

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

Response 5.6:

5.7 DLCs. 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.

Response 5.7: Panel members discussed the need to include DLC-studies in the evaluation and, although panel members appreciated that DLC is a larger question than TCDD alone, several panel members pointed out the scientific importance and regulatory relevance of including a coordinated TEQ/DLC-discussion in the response. Including TEQ/DLC-aspects in the evaluation will open up possibilities to weigh in challenges as well as opportunities. Including TEQ/DLC-aspects will also lead to less risk of overestimating the TCDD-effect (e.g., NIOSH cohort).

5.8 Non-linear approach. 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.

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

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.

Response 5.8a,b: Panel members agreed that the presented non-linear approaches need considerable improvement. This was the point where several panel members felt that the Agency did not respond to the NAS comments.

Section 6. Feasibility of Quantitative Uncertainty Analysis from NAS Evaluation of the 2003 Reassessment (Summary provided by Dr. Ferson)

The following summary has been organized to respect the four charge questions, although the panel's winding discussion did not perfectly do so.

6.1: Clearly presented and scientifically justified?

The EPA response is clearly presented, although one panel member felt the introductory matter was overly pedantic and that the whole section should be rewritten to be accessible by non-statisticians. Some phrasing and words choices in the text should be reconsidered, including 'exotic methods', 'volitional uncertainty', and 'epistemic uncertainty'. One panel member thought the definition of 'quantitative uncertainty analysis' was overly narrow should be expanded to embrace methods other common and useful methods.

The arguments in section 6 are not scientifically justified. Although EPA's decision to not do a quantitative analysis might have been justified on grounds of practicality, the panel feels that quantitative uncertainty analysis is an integral part of any good assessment, and many issues in

this case beg for explicit consideration in the context of an uncertainty analysis. The panel thought that EPA should be methodical and balanced about what variables and components of the assessment would be included in the analysis. The uncertainty narratives and sensitivity analyses already in the document are an excellent beginning and may constitute the lion's share of the work necessary to implement quantitative uncertain analysis based on simple bounding.

6.2: Comprehensive QUA is unfeasible?

The panel rejects EPA's argument that a quantitative uncertainty analysis is unfeasible. Many on the panel felt that the present circumstances warrant a compromise approach that would be simple and achievable with modest effort by the agency. Various bounding approaches, sensitivity studies, and event trees (probability trees without the probabilities) were suggested as possible approaches that could be used. With such methods, legitimate and comprehensive uncertainty analyses (including even probabilistic analyses) are possible. They would be useful and sufficient to respond to NAS' criticism.

Many on the panel further suggested that value-of-information methods would also be very useful, although feedback from EPA included reservations about this idea.

The panel agreed with EPA that expert elicitation would be problematic and should be off the table.

The panel will assemble several issues into a *parking lot*, which arose in its consideration of the various other sections of the EPA document, that should be addressed in the eventual uncertainty analysis.

6.2a: Volitional uncertainty?

Several on the panel bristled at the term 'volitional uncertainty', which might also be called 'decisional uncertainty'.

Several panel members mentioned that standard tools and approaches from decision theory could be used.

Feedback from SAB staff emphasized that EPA's work at this stage is not a "decision" but rather only an assessment, and cautioned against the careless use of the term.

Further feedback from EPA mentioned that decision theory focuses on finding good decisions rather than propagating uncertainty about which decisions will be made.

6.3: Utility of the sensitivity studies?

The panel congratulates EPA on the sensitivity studies that it has already done and considers them to be very useful. The panel felt these studies should be integrated and unified in an overall uncertainty analysis. The panel emphasized that EPA has already done the lion's share of the effort needed already in their considerations described in the uncertainty narratives. The panel

feels the agency should take credit for this hard work and extend them to respond fully to the NAS criticism.

Notes to the Panel

Perhaps the panel can come to consensus that:

- Bounding analysis *is* an uncertainty analysis
- At a minimum, EPA could propagate simple bounds
- There are several ways you could do QUA without expert elicitation
- Probability tree or even just an event tree would be helpful
- Sensitivity studies, even if not completely comprehensive, could be useful
- Epistemic uncertainty is not what the document says it is
- An appropriate QUA is possible, though EPA may decline to do one on other grounds
- Possible bounding approaches include
 - **Interval analysis** (Moore 1966; Neumaier 1990) which has been widely used for decades,
 - **Nesting of intervals** (an approach which philosophers sometimes call “supervaluation” in the sense of van Fraassen) by
 - **Info-gap decision theory** (Ben-Haim 2006) which has been used in several applications (Davidovitch et al. 2009; Hall and Harvey 2009; Regan et al. 2005; Rout et al. 2009; Yokomizo 2009), and
 - **Probability bounds analysis** (Ferson and Long 1995; Ferson 2002; Ferson et al. 2003) including Bayesian p-boxes (Montgomery 2009), which has been used in a variety of applications (Aughenbaugh and Paredis 2007; Dixon 2007; Karanki et al. 2009; Minnery et al. 2009; Regan et al. 2002a; 2002b), including two Superfund assessments (EPA 2007; <<>>).
- Model uncertainty, including uncertainty about dependencies, can also be addressed with bounding approaches
- The word ‘exotic’ should be excised from the document
- Validation, via a ‘reality check’ against the total number of cancers, for instance, would be a good thing. [Need to specify exactly how you’d do a validation study in practice]
- Selecting precise probability distributions may be hard, but ranges are easy
 - [Harvey Clewell suggests focusing on “spread” if not “uncertainty” and wants to put bounds on everything]
 - [Jeffrey Fisher said “Distributions can be used, or ranges if you will”]