



Comments to the EPA Science Advisory Board on the 2010 Dioxin Risk Reassessment

October 18, 2010

These comments are submitted to the US Environmental Protection Agency's Science Advisory Board pursuant to its review of EPA's 2010 Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments (EPA/600/R-10/038A). The comments were prepared by Dr. Gail Charnley of HealthRisk Strategies in Washington, DC and Drs. Lorenz Rhomberg and Robyn Prueitt of Gradient, based in Cambridge, MA and Seattle, WA. HealthRisk Strategies provides independent policy analysis of issues relating to the assessment, management, and regulation of public health risks from chemical exposures. Gradient is an environmental and risk science consulting firm that specializes in employing sound science to resolve complex problems relating to chemicals in the environment, in the workplace, and in consumer products. These comments were prepared at the request of the Research Foundation for Health and Environmental Effects, a 501(c)(3) non-profit organization established by the American Chemistry Council's Chlorine Chemistry Division that supports joint research projects sponsored by industry, public agencies, academia, and other foundations.

Our comments address three areas: weight-of-evidence analysis, risk assessment of cancer effects, and risk assessment of noncancer effects. We are particularly concerned that EPA's 2010 dioxin reassessment fails to follow EPA's own risk assessment guidance as embodied in its 2000 Risk Characterization Handbook,¹ 2002 Information Quality Guidelines,² 2003 Assessment Factors handbook,³ 2004 Risk Assessment Principles and Practices documentation,⁴ and 2005 Guidelines for Carcinogen Risk Assessment,⁵ and that it ignores the recommendations of the National Academy of Sciences committee that reviewed EPA's 2003 dioxin reassessment. The 2010 reassessment does not evaluate or portray the true weight of the scientific evidence and its assumptions about dioxin's carcinogenic mode of action are poorly supported. Its linear dose-response justification would set a precedent as a major science policy departure from accepted practice in the absence of the larger and

¹ EPA (2000) Risk Characterization Handbook. EPA 100-B-00-002. Science Policy Council, Washington, DC

² EPA (2002) Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency. EPA/260R-02-008. Office of Environmental Information, Washington, DC

³ EPA (2003) A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information. EPA 100/B-03/001. Science Policy Council, Washington, DC

⁴ EPA (2004) Risk Assessment Principles and Practices. EPA/100/B-04/001. Office of the Science Advisor, Washington, DC

⁵ EPA (2005) Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F. Risk Assessment Forum, Washington, DC

fuller peer review that would be required for such a departure. The noncancer endpoints used for risk assessment are of questionable clinical relevance.

Thank you for the opportunity to submit these comments. We are happy to provide any additional information upon request.

WEIGHT-OF-EVIDENCE ANALYSIS

(1) The primary shortcoming of EPA's 2010 dioxin reanalysis is that it fails to evaluate the potential human cancer and noncancer effects of dioxin using a weight-of-evidence analysis, despite the direction to do so provided by its own risk assessment guidance documents and by the National Academy of Sciences committee that reviewed EPA's 2003 dioxin reanalysis.

A weight-of-evidence analysis for any potential health effects, including those for cancer or noncancer endpoints, should be more than a matter of describing a set of available studies with an array of results and then announcing one's overall professional judgment. It is important to be systematic and transparent about the information being drawn from the studies, the method used for evaluation and formulation of judgments, and the scientific reasoning behind any judgments offered. EPA's own Risk Characterization Handbook includes criteria for transparency in risk assessment so that any reader can understand all the steps, logic, key assumptions, limitations, and decisions made, and can easily comprehend the supporting rationale that lead to the outcome [p. 15]. Because judgments made about potential risk will usually not be definitive, it is important to present the strengths and weaknesses of alternative judgments that could be made, giving the reader a picture of how strongly one or another interpretation is supported vis-à-vis alternative possible explanations. This process is clearly mandated by EPA's guidance, as documented below. If, in the end, a position is espoused for which other reasonable conclusions could be drawn, and especially if the preferred position is chosen on the basis of a science policy or risk management consideration in the face of scientific uncertainty, it is important to forthrightly document this, rather than simply to present the chosen conclusion with a recitation of its supporting evidence, as EPA has done for both the cancer and noncancer findings in its 2010 dioxin reassessment.

Both the NAS review panel and EPA's own guidance recommend a weight-of-evidence process to evaluate the biological plausibility of potential human health effects. For example, EPA's 2002 Information Quality Guidelines recommend a weight-of-evidence approach in risk assessments.

- In the Agency's development of "influential" scientific risk assessments, we intend to use all relevant information; . . . evaluate that information based on sound scientific practices as described in our risk assessment guidelines and policies; and reach a position based on careful consideration of all such information (i.e., a process typically referred to as the "weight-of-evidence" approach). [p. 26]

Similarly, EPA's 2003 Assessment Factors Handbook addresses the need for weight-of-evidence analysis in risk assessment.

- The weight-of-evidence approach generally considers all relevant information in an integrative assessment that takes into account the kinds of evidence available, the quality and quantity of the evidence, the strengths and limitations associated of each type of evidence, and explains how the various types of evidence fit together. [p. 2]

EPA's 2004 Risk Assessment Principles and Practices document also advises the use of a weight-of-evidence evaluation.

- Risk assessment involves consideration of the weight of evidence provided by all available scientific data . . . Judgment on the weight of evidence involves consideration of the quality and adequacy of data and consistency of responses induced by the stressor. [p. 71]

In particular, a weight-of-evidence process should be used *prior to* the selection of studies for quantitative dose-response analysis, to integrate all relevant information on a particular response in a comprehensive and transparent manner.

The NAS committee that reviewed EPA's 2003 dioxin reassessment recognized the shortcomings of EPA's approach to evaluating potential human health effects and specifically recommended that the Agency perform a weight-of-evidence evaluation for relevant endpoints.

- . . . the committee notes that EPA does not use a rigorous approach for evaluating evidence from studies and the weight of their evidence in the Reassessment. [p. 47]
- The Reassessment provides an extensive catalog of studies but does not synthesize the significant insights or provide clear assessments of the key uncertainties in a way that allows the reader to determine the impact of various choices made. [p. 48]
- [T]he EPA Reassessment . . . relies largely on committee-based, consensus evaluation of the available data rather than on specifically commissioned, rigorous analyses constructed according to established criteria that both formally evaluate the strengths of the available evidence and integrate, by quantitative systematic review, the data across available studies. [pp. 163-164]
- The divergent data across the diverse studies assessing human noncancer end points have not been subjected to systematic review according to currently accepted approaches . . . nor has there been formal grading of the quality of the evidence according to accepted principles . . . [p. 173]
- For available human, clinical, noncancer end point data, EPA should establish formal principles of and a formal mechanism for evidence-based classification and systematic statistical review, including meta-analysis when possible. [p. 174]
- The quality of the available evidence should be reported, and the strength or weakness of a presumptive association should be classified according to currently accepted criteria for levels of evidence. [p. 196]

In the 2010 reanalysis, EPA did not follow the recommendations of the NAS committee, or those of their own guidance, to conduct a weight-of-evidence evaluation of potential effects of dioxin exposure.

Instead, EPA presented their study inclusion criteria and evaluation considerations for both cancer and non-cancer data. More specifically, EPA's study inclusion criteria preclude a weight-of-evidence analysis because they select solely for epidemiologic studies that demonstrate "an association between TCDD and an adverse health effect" [p. 2-7] or for which the "magnitude of animal responses is outside the range of normal variability exhibited by control animals" [p. 2-8]. EPA's inclusion criteria specifically exclude studies that demonstrate no effect, effectively preventing a balanced consideration of available evidence supporting or refuting the biological plausibility and likelihood of effects. Thus, the inclusion criteria relied upon in EPA's 2010 dioxin reassessment specifically violate the recommendations of its own 2002 Information Quality Guidelines, 2003 Assessment Factors Handbook, 2004 Risk Assessment Principles and Practices documentation, and the recommendations of the NAS committee that reviewed the 2003 dioxin reassessment. More generally, EPA's approach violates the criteria for transparency, consistency, and reasonableness found in its 2000 Risk Characterization Handbook.

A true weight-of-evidence analysis should explicitly present the criteria for inclusion and exclusion of studies so that *all* relevant information is included and so that biases toward inclusion of certain outcomes (e.g., only positive outcomes) are avoided. That is, negative or inconsistent results are important to address because their existence will have to be part of the overarching explanation of the array of results on hand. It is important to be explicit about what results are being drawn from each study and not focus just on positive outcomes. Methodologic strengths and weaknesses of each study should be noted without respect to study outcome in order to better assess similarities and differences in study outcomes. The goal is to be able to interpret possible reasons for disagreement, not to select the "best" study and rely on it even if it is contradicted by other study results.

Study results should be arrayed in such a way that does not unduly emphasize positives over negatives and, moreover, that attends to the reasoning and pitfalls involved with deciding what endpoints (and what measures of those endpoints) and what dose measures are to be considered comparable in comparisons across studies. In particular, creating a general category of response and then treating individual studies as corroborative even if the particular responses from study to study differ (though they may be in the same overarching category) can bias the analysis by failing to note the lack of corroboration of particulars. For instance, the assertion that dioxins lead to a broad increase in all human cancers, the particular studies that find increases only in particular cancers, or different studies that find increases in different kinds of cancer from one study to another, are in fact contradictory unless there is evidence of some basis for a general carcinogenic mechanism to act in different particular ways in different settings.

Performing a true weight-of-evidence analysis is consistent with requirements by all three branches of the federal government to use the best available scientific information in order to produce balanced, high quality decisions. For example, President Clinton's Executive Order 12866 (still in force) stipulates that agencies should base their regulatory decisions on the "best reasonably obtainable scientific, technical, economic, and other information."⁶ Congress has consistently underscored a national policy requiring agencies to promulgate science-based regulations. For example, rules promulgated under the Safe Drinking Water Act must use the "best available, peer reviewed science" and present

⁶ Federal Register, Volume 58, No. 190 (October 4, 1993)

“comprehensive, informative, and understandable” risk information. Furthermore, the US Supreme Court’s Daubert decision established that expert opinion based on a scientific technique is inadmissible in lawsuits if the technique is not generally accepted as reliable in the relevant scientific community.⁷ Thus, all branches of the federal government underscore the need to assess all available scientific information. Doing so requires a weight-of-evidence process that is consistent, comprehensive, balanced, and reproducible in risk assessment.

EPA itself addresses the use of best available scientific information in a variety of documents. For example, EPA’s 2002 Information Quality Guidelines define a weight-of-evidence approach and recommends that approach for risk assessment [p.26]. EPA clarified that recommendation in its 2003 Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information, which was intended to assure data transparency and to provide guidance for EPA’s weight-of-evidence analyses. According to that document, such analyses are meant to consider “all relevant information in an integrative assessment that takes into account the kinds of evidence available, the quality and quantity of the evidence, the strengths and limitations associated with each type of evidence and explains how the various types of evidence fit together [p. 2]. Similarly, EPA’s 2004 document Examination of Risk Assessment Principles and Practices makes a commitment to assess all available scientific information using a weight-of-evidence process that is consistent, comprehensive, balanced, and reproducible. Moreover, a weight-of-evidence approach is embraced as a key feature of EPA’s 2005 Guidelines for Carcinogenic Risk Assessment [p.1-11].

Finally, the need for a weight-of-evidence evaluation is at the heart of the recommendations made by the NAS committee that reviewed EPA’s 2003 Draft Dioxin Reassessment. Weight-of-evidence analysis is not a novel concept in EPA’s risk assessment paradigm and is addressed in numerous EPA guidance documents. Absence of a true weight-of-evidence approach from the 2010 Dioxin Reanalysis constitutes a glaring omission in light of these guidelines and policies.

CANCER

(2) EPA’s 2010 dioxin reanalysis states that there is insufficient evidence to support the use of a nonlinear cancer dose-response model, defaulting to a low-dose linear model instead. That conclusion is in conflict with the unanimous conclusions of the National Academy of Sciences review panel, with EPA’s own guidance and procedures, and with virtually every other scientific and regulatory government organization in the world that has reviewed dioxin.

Instead of following the recommendations of the National Academy of Sciences and, in conflict with its own cancer risk guidelines, EPA’s 2010 dioxin reassessment continues to rely on a linear model for TCDD, adding some nonlinear calculations only as “illustrative examples”. There is no balanced weight-of-evidence analysis of the science supporting linearity versus nonlinearity and the reassessment reads like a lengthy justification for the predetermined policy choice of linearity.

⁷ Daubert v. Merrell Dow Pharmaceuticals Inc., 516 U.S. 869 (1993)

The 2010 reassessment's justification for choosing linearity is that TCDD's carcinogenic mode of action is unknown.

- The sequence of key events following binding of TCDD to the AhR and that ultimately leads to the development of cancer is unknown. [pp. 5-10 to 5-11]
- The mode of action of TCDD in producing liver cancer in rodents has not been elucidated. [p. 5-17]
- . . . a defined mechanism at the molecular level or a defined mode of action for TCDD-induced carcinogenicity is lacking . . . [p. 5-20]
- EPA believes that the mode of action is not known, so is using the default linear extrapolation approach specified by EPA's cancer guidelines. [p. 5-63]

In contrast, EPA's cancer guidelines actually state, "At least some information bearing on mode of action . . . is present for most agents undergoing assessment of carcinogenicity, even though certainty about exact molecular mechanisms may be rare" [pp. 2-36 to 2-37]. TCDD's exact *mechanism* of action may not be entirely clear, but its *mode* of action is. In fact, the reanalysis notes that the cancer guidelines define mode of action as "a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation" where a "key event" is an empirically observable precursor step that is itself a *necessary element* of the mode of action or is a biologically based marker for such an element [p. 5-10].

The reanalysis acknowledges that the necessary element associated with TCDD's carcinogenic mode of action is AhR receptor-mediated. Receptor-mediated modes of action are generally associated with nonlinear dose-response relationships.⁸

- Most evidence suggests that the majority of toxic effects of TCDD are mediated by interaction with the AhR. EPA considers interaction with the AhR to be a necessary, but not sufficient, event in TCDD carcinogenesis. [p. 5-10]

Furthermore, in its discussion of the plausibility of TCDD-induced human carcinogenesis, the reanalysis refers to the AhR-mediated mode of action in rodents.

- Several hypothesized modes of action have been presented for TCDD-induced tumors in rodents, all involving AhR activation. The available evidence does not preclude the relevance of these hypothesized modes of action to humans. [p. 5-9]
- TCDD is characterized as carcinogenic to humans [based on] general scientific consensus that the mode of TCDD's carcinogenic action in animals involves AhR-dependent key precursor events . . . [p. 5-20]

⁸ See, e.g.: NAS/NRC (2006), Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment; Ross and Kenalkin (2001), Pharmacokinetics: Mechanisms of drug action and the relationship between drug concentration and effect, pp. 31-43 in Goodman & Gilman's the Pharmacological Basis of Therapeutics, 10th Ed.; Kohn and Melnick (2002) J. Mol. Endocrinol. 29:113

Then, reiterating the assertion that TCDD's mode of action is unknown, the reanalysis chooses the low-dose-linear model as the appropriate default model for describing TCDD's dose-response. However, EPA's cancer guidelines explicitly state that both linear and nonlinear dose-response models can be considered "default" approaches.

- [D]efault approaches can be applied that are consistent with current understanding of mode(s) of action of the agent, including approaches that assume linearity or nonlinearity of the dose-response relationship, or both. [p. 1-14]

The cancer risk guidelines do *not* require full understanding of a nonlinear mode of action to support a nonlinear dose-response model, as long as there is significant scientific support for nonlinearity.

- Nonlinear extrapolation having a significant biological support may be presented in addition to a linear approach when the available data and a weight of evidence evaluation support a nonlinear approach, but the data are not strong enough to ascertain the mode of action applying the Agency's mode of action framework. [p.3-23]

If no scientific consensus exists regarding mode of action, the results of both linear and nonlinear approaches are shown.

- Where . . . no scientific consensus favors a single approach, an assessment may present results using alternative approaches. A nonlinear approach can be used to develop a reference dose or a reference concentration. [p. 1-15]

The decision about which approach is most appropriate then becomes a risk-management decision.

- When risk assessments are performed using only one set of procedures, it may be difficult for risk managers to determine how much health protectiveness is built into a particular hazard determination or risk characterization. When there are alternative procedures having significant biological support, the Agency encourages assessments to be performed using these alternative procedures, if feasible, in order to shed light on the uncertainties in the assessment, recognizing that the Agency may decide to give greater weight to one set of procedures than another in a specific assessment or management decision. [p. 1-8]

The cancer guidelines also state that a decision about a substance's carcinogenic mode of action should reflect current scientific understanding, where "current understanding" [p. 1-14] of an agent's mode of action is to be determined based on a weight-of-evidence analysis.

- All pertinent studies are reviewed in analyzing a mode of action, and an overall weighing of evidence is performed, laying out the strengths, weaknesses, and uncertainties of the case as well as potential alternative positions and rationales. [p. 2-41]

However, what the reanalysis describes as its weight-of-evidence analysis [p. 5-3ff] is, in fact, a summary of the evidence EPA believes supports its classification of TCDD as carcinogenic to humans, not a weight-of-evidence analysis. Excluding studies that do not demonstrate a dose-response provides an

unbalanced context for those studies that do, and eliminates from consideration studies that provide useful information for understanding the range of uncertainty. Furthermore, omitting endpoints or studies that do not show a dose-response relationship in the direction EPA expects may discount valuable information, particularly information that could inform mode of action as well as dose-response.

According to the cancer guidelines, a decision about a substance's carcinogenic mode of action should also reflect current scientific understanding by determining the extent to which scientific consensus generally supports a particular mode of action.

- In reaching conclusions, the question of “general acceptance” of a mode of action should be tested as part of the independent peer review that EPA obtains for its assessment and conclusions. [p. 2-40]

The concept of “general acceptance” is also reflected by the reasonableness criteria specified in EPA's Risk Characterization Handbook.

- Reasonableness . . . demonstrates that the risk assessment process followed an acceptable, overt logic path and retained common sense in applying relevant guidance. [p. 18]
- Reasonableness is achieved when the risk characterization is determined to be sound by the scientific community . . . [and] . . . the assessment uses generally accepted scientific knowledge . . . [p. 18]

The question of “general acceptance” of TCDD's mode of action and choice of dose-response model is one that was put to the National Academy of Sciences committee that reviewed EPA's 2003 dioxin assessment. The committee was asked to evaluate “the validity of the nonthreshold linear dose-response model and the cancer slope factor calculated by EPA through the use of this model” [p. xvi]. The committee concluded unanimously that relying on a linear dose-response model for TCDD is not supported scientifically and that the weight of evidence supports nonlinearity.

- The committee concludes that EPA's decision to rely solely on a default linear model lacked adequate scientific support. [p. 5]
- . . . the committee unanimously agreed that the current weight of scientific evidence on the carcinogenicity of dioxin is adequate to justify the use of nonlinear methods to extrapolate below the [point of departure]. [p. 16]
- The committee concludes that EPA did not support its decision adequately to rely solely on this default linear model . . . The committee determined that the available data support the use of a nonlinear model, which is consistent with receptor-mediated responses and a potential threshold . . . [p. 24]
- . . . the committee concludes that, although it is not possible to scientifically prove the absence of linearity at low doses, the scientific evidence, based largely on mode of action, is adequate to favor the use of a nonlinear model that would include a threshold response over the use of the default linear assumption. [p. 122]

- There is general consensus in the scientific community that nongenotoxic carcinogens that act as tumor promoters exhibit nonlinear dose-response relationships, and that thresholds (doses below which the expected response would be zero) are likely to be present. [p. 122]
- The committee unanimously agrees that the current weight of evidence on TCDD, other dioxins, and [dioxin-like compounds] carcinogenicity favors the use of nonlinear methods for extrapolation below the point of departure (POD) of mathematically modeled human or animal data. [p. 190]
- Quantitative evidence of nonlinearity below the point of departure (POD), the ED01 (effective dose), will never be available because the POD is chosen to be at the bottom end of the available dose-response data . . . EPA should give greater weight to knowledge about the mode of action and its impact on the shape of the dose-response relationship. The committee considers that the absence of evidence that argues against linearity is not sufficient justification for adopting linear extrapolation, even over a dose range of one to two orders of magnitude or to the assumption of linearity through zero, which would not normally be applied to receptor-mediated effects. [p. 178]

However, in concluding that a linear dose-response could not be completely ruled out, the committee recommended that, consistent with EPA's cancer guidelines, EPA's assessment of dioxin should present both linear and nonlinear models accompanied by a balanced description of the weight of evidence supporting each approach, all of which would communicate uncertainty better to the risk manager.

- The report recommends that EPA provide risk estimates using both nonlinear and linear methods to extrapolate below [points of departure]. [p. 5]
- The committee recommends adopting both linear and nonlinear methods of risk characterization to account for the uncertainty of dose-response curve shape below ED01. [p. 72]
- . . . the committee recognizes that it is not scientifically possible to exclude totally a linear response at doses below the POD, so it recommends that EPA provide risk estimates using both approaches and describing their scientific strengths and weaknesses to inform risk managers of the importance of choosing a linear vs. nonlinear method of extrapolation.

Thus, while believing that the science supports the choice of a nonlinear dose-response model over a linear model for TCDD, the National Academy of Sciences committee that reviewed EPA's 2003 dioxin assessment recommended that EPA provide results using both modeling approaches, accompanied by a discussion of the strengths and weaknesses of each, so that the extent of the uncertainty would be transparent. Although the committee did not believe that the science supported linearity, it recognized that completely ruling out low-dose linearity would never be possible scientifically and that "[t]o the extent that EPA favors using default assumptions for regulating dioxin as though it were a linear carcinogen, such a conclusion should be made as part of risk management" [p. 190].

The question of "general acceptance" of TCDD's mode of action and choice of dose-response model can also be addressed by comparing EPA's dioxin reassessment to the risk assessments performed internationally by other public health organizations. For example, the World Health Organization states

that “TCDD does not affect genetic material and there is a level of exposure below which cancer risk would be negligible” and that “[t]he experts concluded that a tolerable intake could be established for dioxins on the basis of the assumption that there is a threshold for all effects, including cancer.”⁹ The WHO tolerable daily intake (or some version thereof) has been adopted by most other countries of the world. In addition, the International Agency for Research on Cancer recently noted that TCDD was the first substance to be classified as a known human carcinogen based primarily on sufficient data in animals on both carcinogenicity and mechanism of action, specifically, “sufficient evidence . . . for a mechanism via initial binding to the aryl hydrocarbon receptor (AhR), which leads to changes in gene expression, cell replication, and apoptosis.”¹⁰ EPA’s own Risk Characterization Handbook specifies consistency criteria requiring EPA to include comparisons to assessments done by other agencies and organizations in order to put its own risk assessments in context. Thus in concluding that there are insufficient data with regard to TCDD’s carcinogenic mode of action to justify nonlinearity, EPA’s 2010 dioxin reassessment contradicts its own guidance as well as the generally accepted conclusions of esteemed international scientific organizations.

(3) Invoking additivity-to-background and population heterogeneity arguments in support of low-dose linearity is a novel application of a new science-policy principle, and should not be done without thorough discussion and peer review.

The arguments about population heterogeneity and the nature and existence of an additivity-to-background effect are complex and use of those arguments as the basis for determining appropriate dose-response analyses has not been widely accepted nor even widely discussed in the scientific community.¹¹ Its use would be a novel and inappropriate application of a new science-policy principle. This should not be done without thorough discussion and peer review. The few brief discussions in EPA’s 2010 reassessment of how the additivity-to-background argument is being invoked for dioxins are insufficient to provide a basis for such a major science policy departure from accepted practice. EPA’s argument for linearity should not be accepted without a larger and fuller review as an element of science policy.

The reassessment states that there is insufficient information to establish a threshold for dioxin-induced carcinogenesis because, although a particular receptor-mediated event in an individual may have a threshold, there will be a distribution of thresholds at the population level that may or may not bear a resemblance to an individual’s receptor kinetics.

- . . . in general, the population dose-response curve depends on (1) the distribution of individual thresholds in the neighborhood of zero, (2) the dose-response curve for each individual, and (3) the dose metric. Under EPA’s Cancer Guidelines, the zero-slope-at-zero criterion applies strictly to ingested dose, but the other two factors (distribution of individual thresholds and dose-

⁹ <http://www.who.int/mediacentre/factsheets/fs225/en/index.html>

¹⁰ Baan et al. (2009), www.thelancet.com 10:1143

¹¹ Rhomberg (2009) *Environ. Health Perspect.* 117:141

response curve for each individual) need to be established before a zero slope at zero dose can be established. Otherwise the default linear extrapolation to zero approach applies. [p. 5-57]

- On the nature or the distribution of individual thresholds, often referred to as the population tolerance distribution, there is ongoing debate as to how receptor kinetics influence the shape of that distribution. Even within an individual, there is a lack of consensus as to whether receptor kinetics confer linear or sublinear attributes to downstream events, or whether receptor kinetics, themselves, are linear, sublinear, or supralinear. Whatever the nature of the form of receptor kinetics, it may have little or no influence on the ultimate population response. [p. 5-57]
- There is no *a priori* reason to believe that the shape of the dose-response curve in an individual has any relationship to the shape of the population response, particularly for quantal endpoints. [p. 5-57]

The reanalysis specifically invokes the additivity-to-background argument in support of low-dose linearity, justifying this argument by referring to a “state-of-the-science workshop” on issues in low-dose risk extrapolation held by EPA and Johns Hopkins Risk Science and Public Policy Institute in 2007¹² and to the 2009 National Academy of Sciences report *Science and Decisions: Advancing Risk Assessment*.

In invoking the additivity-to-background/nonthreshold argument, EPA suggests that endogenous AhR activity provides sufficient induction of gene expression and other down-stream effects – which are presumed to be the same as those induced by dioxins and responsible for high-dose tumorigenicity of dioxins – that even in unexposed populations, some tumors will result from the normal level of operation of such processes. (This is the “background” to which dioxins are being presumed to add.) Exposure to dioxins, in this view, can exacerbate the operation of these processes by providing additional binding to AhR, consequent increased levels of gene expression, consequent increases in down-stream consequences of those expression changes, and hence added risk of tumors by enhancing the magnitude of the process responsible for the background tumors. (This is the additivity effect to the inherent background that is being proposed.)

This schema is a rather specific mode-of-action assertion, requiring acceptance of a whole suite of presumptions about the nature of the tumorigenic process, its operation in the absence of dioxin exposure, and the dose-response relations among a series of intermediate stages. Elsewhere in its document, EPA has asserted that TCDD’s mode of action cannot be determined with sufficient certainty to form the basis for choosing a dose-response curve shape (see discussion of comment #2 above). Therefore, the speculation about TCDD’s mode of action entailed in invoking the additivity-to-background principle is illogical, inconsistent with those other assertions, and unacceptable.

In order for the proposed additivity to work, it must be the case that background tumors result from the same set of failures of control of cell division and differentiation that are induced by dioxins at higher exposures, but there is no basis to assert this.

¹² White (2009) Environ. Health Perspect. 117:283

It must also be the case that individual thresholds exist, such that the discrete event of induction of a new tumor does not happen in all individuals with any level of endogenous AhR activity (because the whole population has such activity, and yet tumors are rare). Moreover, a small increase in an individual's level of AhR activity (as is presumed by the argument to be induced by a small dioxin exposure) must be sufficient to move that person from being a non-responder to having a tumor induced. It must be presumed that endogenous AhR ligands do not act as antagonists to, and therefore inhibitors of, dioxin binding or its efficacy, that displacement of endogenous ligands by dioxins does not simply lead to similar receptor occupancy by different ligands, and that the array of downstream effects of binding of exogenous dioxins and endogenous ligands are the same.¹³ The existence of a dose-response relationship in the population must then be attributed to inter-individual variation in the individual thresholds, and the pattern of this must be such that some individuals have thresholds so low that they respond even without any dioxin exposure (and hence constitute the background), while many others hover on the verge of this level and require only a small dioxin exposure to push them over their individual thresholds. No basis for such a schema is presented and it is difficult to imagine one.

The schema presumes without evidence that any amount of change in the degree of AhR occupancy increases the magnitude of the downstream subsequent processes involved in tumorigenesis without a threshold. It is only in this way that small changes in AhR occupancy can lead to a tumor increase. Yet it is evident from AhR's role in such gene-expression effects as EROD activity that there is nonlinearity and indeed thresholds between the degree of receptor occupancy and the effects induced. This is also observed in most other receptor-mediated processes; the nonlinearity in response and the existence of thresholds comes not from the degree of receptor occupancy, but rather from the interactions of processes (including homeostasis perturbation, positive and negative feedbacks, etc.) downstream to the level of changes in gene expression. With all receptor-mediated processes, it is the complex interaction of such control networks, and not the linearity or nonlinearity of a single component, that dictates the dose-response relationship for the apical effect. The linearity of one component early in the sequence gives little information about this larger behavior of the system. Additivity to AhR occupancy, as invoked by the EPA, does not lead to linearity of these downstream processes. Assuming that all the downstream processes are individually linear and that the outcome of their interaction is linear – which is necessary in order to use linear effects of AhR-binding as evidence for linearity of cancer risk – constitutes assuming the truth of the proposition (dose-response linearity) that one is seeking to explain.

The additivity-to-background argument is an argument in principle, but it does not itself provide any basis for estimating the size of any low-dose linear component, for determining the range of doses over which additivity produces linearity, or whether the effect (even if it exists) substantially alters the dose-response relationship that would be estimated without reference to additivity-to-background.

In particular, even if an additivity-to-background effect occurs, it does not lead to linearity of the whole dose-response curve, but would only affect very small risks at very small doses, with the shape of most of the full dose-response curve (including that part we are able to observe in actual data)

¹³ Safe (1998) *J. Animal Science* 76:134

determined by mode of action. Simply invoking a linear extrapolation from some point higher in the dose-response relationship is not a way to incorporate additivity to background into the analysis. Forcing a linear curve fit or linearly extrapolating from some observable point on the curve results in a measure of low-dose linearity that has nothing to do with the reasons the linearity was invoked, and so such methods do not provide a basis for judging the actual magnitude of a low-dose linear component nor do they address for what limited range of low dose levels and low risk levels the presumed linear relationship should hold before it is overwhelmed by mode-of-action-driven influences on the dose-response shape at more substantial doses that may be of interest to risk assessors. Using such methods will produce misleading and unreliable estimates – most likely radical overestimates – of the actual effect even if the presumptions of the additivity-to-background effect are true.

In conclusion, EPA's decision to apply the new science policy principles of additivity to background and population heterogeneity runs counter to any semblance of a weight-of-evidence perspective and analysis as well as the spirit and intent of EPA's Risk Characterization Handbook. The Handbook's principles state that "[a] risk characterization should be prepared in a manner that is clear, transparent, reasonable, and consistent with other risk characterizations of similar scope prepared across programs in the Agency" [p. 14]. The policy goes on to state that the principles of transparency, clarity, consistency, and reasonableness need to be fully applied throughout every aspect of the risk assessment process.

NONCANCER EFFECTS

(4) EPA's 2010 dioxin reanalysis ignores the recommendations of the NAS review panel and its own guidance by failing to evaluate the clinical relevance of the effects considered for RfD derivation.

The NAS committee that reviewed EPA's 2003 dioxin risk assessment recommended that EPA evaluate the biological relevance of reported effects.

- Attention should also be directed to addressing the potential biological significance of very small statistically significant physiological or biochemical changes that remain well within the normal range of variation and adaptation. [p. 163]

In addition, EPA's 2004 Risk Assessment Principles and Practices document indicates the need to determine the biological relevance of an effect.

- As a general principle, our practice is not to base risk assessments on adaptive, non-adverse, or beneficial events. [p. 53]

In the 2010 reassessment, EPA considered the toxicological relevance of endpoints from animal studies, but did not do the same for human endpoints.

- In selecting POD candidates from the animal bioassays for derivation of the candidate RfDs, EPA had to consider the toxicological relevance of the identified endpoint(s) from any given study.

Some endpoints/effects may be sensitive, but lack general toxicological significance due to not being clearly adverse...being an adaptive response or not being clearly linked to downstream functional or pathological alterations. [p. 4-7]

For humans, EPA provided a brief justification for the use of the two endpoints (elevated TSH levels in neonates and decreased sperm concentration in adult males exposed during childhood) for dose-response modeling, choosing the endpoint with the lowest LOAEL (sperm concentration) for derivation of the RfD. The NAS committee had previously noted that the consideration of these endpoints as “adverse” is highly questionable, and recommended that EPA include a discussion of the magnitude of reported changes and whether they are within the normal range.

- [Regarding elevated TSH levels in the study by Pavuk et al. (2003),] [t]he discussion does not address the fact that the TSH differences, although statistically significant, are quantitatively extremely small and well within the normal range of circulating TSH levels. [p. 170]
- The draft Reassessment also highlights the higher TSH values reported in human infants by Pluim et al. (1993) and by Koopman-Esseboom et al. (1994)...but does not discuss the fact that the TSH changes were very small and possibly not of physiological or clinical significance. [p. 171]
- [Regarding studies of dioxin exposure and reproductive and developmental outcomes,] [t]he committee agrees that the results are subtle but disagrees that the reported effects are truly clinically adverse, especially when confidence in the observations is low and the reported changes could be non-significant at the biological level and clinical outcome. [p. 164]

Overall, the NAS committee concluded that the evidence for dioxin exposure as a cause of reproductive and hormonal abnormalities is not strong.

- Although the spectrum of reported human reproductive and hormonal abnormalities following dioxin exposure is generally similar to that found in animals, the strengths of the individual associations in studies thus far, are weak, and confidence in the causal nature of these associations while suggestive is not compelling. [p. 162]

In fact, the NAS committee stated that there is no convincing evidence of adverse, non-cancer effects as a result of dioxin exposure.

- In humans, the association of TCDD exposure with other reported, detrimental non-cancer effects has not been convincingly demonstrated. The available studies have not yet shown clear associations among TCDD exposures and the risks of individual, clinically significant, non-cancer end points. [p. 173]

Despite those conclusions of the NAS panel reviewing EPA’s 2003 dioxin reassessment, the 2010 dioxin reassessment nonetheless uses these endpoints as a basis for dose-response modeling and derivation of a non-cancer RfD.

For elevated neonatal TSH levels, as reported in the study by Baccarelli et al. (2008), EPA’s 2010 reassessment cites the World Health Organization (WHO) screening value for neonatal TSH concentration as justification for the use of this endpoint.

- The World Health Organization (WHO, 1994) established the 5 $\mu\text{U}/\text{mL}$ standard as an indicator of potential iodine deficiency and potential thyroid problems in neonates. Increased TSH levels are indicative of decreased thyroid hormone (T4 and/or T3) levels. The 5 $\mu\text{U}/\text{mL}$ “cutoff” for TSH measurements in neonates was recommended by WHO (1994) for use in population surveillance programs as an indicator of iodine deficiency disease (IDD). [p. 4-24]

EPA does not discuss whether a neonatal TSH concentration in excess of the WHO screening level of 5 $\mu\text{U}/\text{mL}$ is indicative of an adverse effect nor whether the “elevated” TSH levels of the subjects in the Baccarelli et al. (2008) study fall within the reported reference range for neonatal TSH levels. Neonatal TSH levels vary considerably during the first 24 hours of birth, with a surge of TSH common (and clinically irrelevant) during the first 12 hours of birth. EPA provides no discussion of whether the reported effect is clinically adverse or within the normal range of adaptive responses.

The justification given by EPA in the 2010 reassessment for using the endpoint of decreased sperm concentration, as reported by Moccarelli et al. (2008), also acknowledges reliance on a screening value intended to indicate that further investigation is appropriate, not that an adverse effect is occurring.

- Although a decrease in sperm concentration of 20% likely would not have clinical significance for an individual EPA's concern with the reported decreases in sperm concentration and total number of motile sperm (relative to the comparison group) is that such decreases associated with TCDD exposures could lead to shifts in the distributions of these measures in the general population. Such shifts could result in decreased fertility in men at the low end of these population distributions. While there is no clear cut-off indicating male fertility problems for either of these measured effects, a sperm concentration of 20 million/ml is typically used as a cut-off by clinicians to indicate follow-up for potential reproductive impact in affected individuals. [p. 4-26]

EPA acknowledged that the mean values for sperm concentration in the Moccarelli et al. (2008) study did not fall below the clinical level of concern (20 million/mL), but did not discuss whether there are any actual data to verify that men potentially at the low end of the distribution of sperm concentration values had higher dioxin exposures.

Both the Baccarelli et al. (2008) and Moccarelli et al. (2008) studies describe outcome measures that are useful clinical markers to guide further investigation but are not indicative of adverse effects in and of themselves. EPA does not accompany the use of the data from these studies for dose-response modeling and RfD derivation with a discussion of the clinical significance of the effects or the levels of change that represent an adverse effect for each of the endpoints.

The 2010 reassessment's focus on including data sets based on the simple ability to be subjected to dose-response analysis is a valid consideration, but it should come as the last of a series of considerations. The first consideration should be to establish that the endpoint in question is a valid potential human endpoint. Such a hazard characterization should include a weight-of-evidence analysis across available studies that examines whether the alleged effect is repeatable within settings and generalizable across settings (e.g., to other species), and evaluates what is known about the relevance to humans of the apparent mode of action. An approximate concordance across studies of apparent

effective dose levels, dose timing, and sensitive periods is an important part of establishing the existence of a commonality of causation that might apply to humans.

Once an endpoint is judged to be sufficiently robustly demonstrated and sufficiently plausibly applicable to human exposures, then the analysis should focus on identifying those studies among the set available that are deemed to best represent or exemplify this generally operating causal process. Only then, once this subset of data sets is identified, should the amenability to dose-response analysis enter the consideration, for only among such studies will the results of such an analysis be truly informative about potential human risk. It is important to attend to the measures of response and the arguments about how much change is being considered to be necessary for a relevantly adverse impact.

For example, a recent weight-of-evidence analysis for dioxin and non-cancer effects showed that there are no substantial, consistent effects of dioxins on thyroid endpoints in infants and children (Goodman et al., 2010). This evaluation looked for consistency and patterns within and across studies and examined whether associations were real and reproducible. The use of this type of rigorous analysis for all potential effects allows for the identification of endpoints with the strongest evidence for causality. Based on a weight-of-evidence review such as this, key studies for the endpoint(s) that will be considered in a subsequent quantitative dose-response analysis can be chosen with greater confidence in their relevance.