External Peer Review

Toxicological Review and IRIS Summary for 1,4-Dichlorobenzene (Inhalation Cancer Assessment and RfC)

FINAL REPORT

Prepared for
Integrated Risk Information System (IRIS) Program
Office of Research and Development
National Center for Environmental Assessment
U.S. Environmental Protection Agency

Prepared by
ORISE IRIS Technical Assistance Team
Oak Ridge Institute for Science and Education
Oak Ridge Associated Universities

November 2006

This document was prepared for the EPA by ORISE under Interagency Agreement No. DW-89939822-01-0 between EPA and the U.S. Department of Energy. ORISE is managed by Oak Ridge Associated Universities under a contract with DOE.

EXTERNAL PEER REVIEWERS

Rogene F. Henderson, PhD, DABT, Chair
Lovelace Respiratory Research Institute
Albuquerque, NM

Bruce C. Allen
Chapel Hill, NC

Abigail Jacobs Ph.D.
Center for Drug Evaluation and Research (CDER)/FDA
Washington, DC

The ORISE IRIS Technical Assistance Team has neither altered nor edited these comments for grammatical or other errors.

PEER REVIEW PROJECT

Audrey Galizia, Project Manager
U. S. Environmental Protection Agency

Leslie Shapard, Senior Peer Review Manager
Oak Ridge Institute for Science and Education

1,4-Dichlorobenzene Review

TABLE OF CONTENTS

REVIEWER COMMENTS BY QUESTION

Derivation of the RfC for 1,4-DCB

Question I 1

Question II 2

Question III 4

Question IV 4

Question V 6

Inhalation Cancer Assessment for 1,4-DCB

Question I 7

Question II 8

Question III 11

Question IV 12

Question V 13

External Peer Review of the
Toxicological Review and IRIS Summary for 1,4-Dichlorobenzene
(Inhalation Cancer Assessment and RfC)

The Toxicological Review of Dichlorobenzenes (DCBs), including the 1,2-, 1,3-, and 1,4-isomers, was subject to an external peer review in February 2004. Revisions to the health assessment were made in response to external peer review comments, which are summarized in Appendix A of the Toxicological Review. The February 2004 external peer review identified a 2-year inhalation toxicity study of 1,4-DCB (Japan Bioassay Research Center [JBRC], 1995) that was subsequently published in the peer-reviewed literature by Aiso et al. (2005). The JBRC study was used as the basis for deriving a revised inhalation reference concentration (RfC) and an inhalation cancer assessment for 1,4-DCB.

The scope of the current external peer review is limited to analyses of dose-response data from the JBRC inhalation study of 1,4-DCB that were not included in the February 2004 external peer review draft. The charge questions below specifically address the revised inhalation RfC and the inhalation cancer assessment for 1,4-DCB.

Derivation of the RfC for 1,4-DCB

1. The principal study used to derive the RfC is the JBRC 2-year inhalation bioassay (Aiso et al., 2005). Is this study the most appropriate selection for the principal study (i.e., the best study upon which to determine the point of departure)? Has the rationale for this choice been transparently and objectively described?

Rogene F. Henderson, PhD, DABT, Chair

I agree that the Aiso 2-year study is the most appropriate for derivation of the RfC. The study was well designed and was done under GLP conditions. It is a chronic study, which is appropriate for derivation of an RfC. The reasons for the choice are well described and objective.

Bruce C. Allen

It is my belief that the JBRC study is the most appropriate selection for the principal study. It is the only full-lifetime inhalation study that was reported in the toxicological review. As such, it is the best basis for deriving risks associated with chronic exposures to 1,4-DCB.

However, the rationale for this choice appears to be missing from the toxicology review. It is in neither section 4.5.2.3, which is intended to present the synthesis and evaluation of major noncancer effects for inhalation of 1,4-DCB, nor section 5.2.3.1, which is where the principal study and critical effect – "with rationale and justification" – are to be presented. Nor is that rationale presented in the proposed IRIS entry. I did not see a discussion of the merits of this study as compared to the other studies that were available. I do not think that there needs to be much discussion associated with this issue; as stated above, it is really the only full-lifetime inhalation study available. But as the documents stand now, a reader would not be able to conclude that that rationale (and the associated considerations of quality and completeness of the

testing and reporting) was the reason for selecting the JBRC study. This is particularly important because there was a previous RfC developed, and there is no justification that I have seen that suggests why the new one should be preferred to the old one.

I would recommend that a more complete discussion of the rationale be provided, probably in section 5.2.3.1. I do not believe, as will be discussed below, that simply citing this study as the one that gave the lowest LOAEL or NOAEL is sufficient in this context.

Abigail Jacobs Ph.D.

The most appropriate study-an inhalation study- was used for evaluation of respiratory tract effects. Oral studies might be used for systemic effects if metabolism is not radically different by the two routes.

Rationale is given and described. However discussion of why eosinophilic changes in olfactory epithelium and testicular mineralization are considered critical effects is needed.

a. Since the nasal passage of the rats and humans are quite different, and rats have a proportionally much larger olfactory epithelium, and the fact that the same lesion was seen in so many of the female rat controls at moderate or marked severity suggests that these rats have an unusual susceptibility and that quantitative extrapolations will be shaky.

b. It is not clear that mineralization of the testes is even present, since supplemental information said that the effect was in the vessels and not the testicular cells. Furthermore, the fact that the same lesion was seen in 55% of the control male mice in this study and in none of the B6C3F1 mice in the NTP study, suggests that these mice have an unusual susceptibility and quantitative extrapolations of any kind will be shaky.

2. The endpoints considered as possible critical effects were eosinophilic changes of the olfactory epithelium in rats and mineralization of the testes in male mice. Please comment on the biological significance of these two endpoints. Was the most appropriate critical effect (eosinophilic changes of the olfactory epithelium) selected as the critical effect? Has the rationale and justification for selection of this critical effect been transparently described? Is the selection of the critical effect scientifically justified?

Rogene F. Henderson, PhD, DABT, Chair

The nasal lesions in the form of accumulation of eosinophilic globules in the cells of the olfactory epithelium are indicative of degenerative changes and thus are appropriate adverse effects to use as an endpoint. However, it is not certain how relevant to humans this endpoint is. The mineralization of the testes does not seem appropriate as an endpoint. A review of the slides by a pathologist commenter (Dr. Creasy) indicated that the lesions was actually in the blood vessels and not in the testis itself. Also, the lesion is typical of an aging process and had a high incidence in controls. Therefore the “testicular” lesion does not seem appropriate and it is good that it was not used. The description of the lesion in the IRIS document should be modified in light of Dr. Creasy’s comments.

Bruce C. Allen

Not being a toxicologist or biologist, I am perhaps not the best person for commenting on the biological significance of the possible critical endpoints. However, I was struck by the fact that testes mineralization in mice was not an endpoint that had been reported in other studies nor did it appear to be discussed in such a way so as to suggest that its occurrence in the JBRC study was consistent with the largely renal or hepatic effects seen in other studies or the respiratory tract effects suggested by the limited human data.

In fact, several of the public comments received recently suggest that neither eosinophilic changes nor tested mineralization are appropriate endpoints because they are not really indicative of adverse changes. In light of those comments, I believe that the document needs much greater attention to the reasoning behind the choice of endpoints for RfC determination.

In the case of the testes mineralization, it appears that the sole basis for its selection was that is was the "most sensitive" endpoint for that study and that species. Even then, I think that that criterion was not necessarily reasonably applied. The basis for that statement is the following. If a dose-response modeling (e.g., BMD) approach is to be the basis for estimating a point of departure, then that approach should be used also to define the critical effects. It is entirely possible that an effect that does not give the lowest NOAEL (and therefore might not be considered as a sensitive endpoint from a NOAEL/LOAEL perspective) could give a lower BMD/BMDL than the endpoint with the lowest NOAEL. This is because the entire curve shape is considered in a dose-response approach, unlike the NOAEL/LOAEL approach that considers only the effect at discrete points (the experimental dose levels).

That "reversal" of the critical endpoint given the adoption of a dose-response modeling approach can happen within a study, but is even more likely to occur when different studies (with perhaps different experimental doses) are compared. Adding to the potential for missing the real critical study is the fact that different HEC conversions may take place for inhalation exposures, even within study, depending on where the effects occur. All of these factors make it difficult, if not impossible, to pick the critical endpoint solely by looking at the NOAEL/LOAEL values and picking the endpoints that give the lowest of those.

I have done a little bit of analysis of one other mouse effect reported in the review (centrilobular hepatocellular hypertrophy in males, with a dose-response of 0/49, 0/40, 0/50, and 34/49). If this were the only other effect in mice that showed some effect due to 1,4-DCB exposure, then the selection of the testes mineralization can be justified (assuming the same HEC conversion since it is an effect distant from the respiratory tract), on the basis of providing the most sensitive dose-response. Thus, my comments above are not meant to suggest that the critical endpoints for the JBRC study were incorrectly identified, but rather that the justification (from a purely regulatory/modeling perspective as opposed to biological/toxicological reasonableness) needs to be cast in other terms consistent with the modeling approach that, rightly, is the basis for the estimation of the point of departure.

Abigail Jacobs Ph.D.

Biologic significance of endpoints
a. eosinophilic changes in olfactory epithelium in rats: (See also no. 1 above.)

b. mineralization of testes- This was not described in the published study and supplemental information said that the effect was in the vessels and not the testicular cells. (See also no. 1 above.)

3. Inhalation dosimetry methods were used in the calculation of the human equivalent concentration (HEC) based on U.S. EPA (1994). Is the explanation for dosimetry choices in the derivation of the RfC scientifically justified and transparently described?

Rogene F. Henderson, PhD, DABT, Chair

I went through the calculations and they seemed scientifically justified and transparent. Using the ratio of minute volume to the surface area affected was a logical approach to estimate dose to tissue.

Bruce C. Allen

In the absence of recapitulating the entire justification (from EPA, 1994) for treating category 2 gases as category 1 for respiratory effects and category 3 for extra-respiratory effects, there is not much more that can be added to scientifically justify the HEC conversions. That does detract from transparency, perhaps, but interested readers should seek out the referenced methodology. What might be useful in that regard, however, is a somewhat more complete rationalization of why 1,4-DCB should be considered a category 2 gas (i.e., its properties with respect to reactivity and solubility that confirm such a categorization, as opposed to just noting that effects were observed both in the respiratory tracts and at remote sites). In fact, the proposed IRIS entry does not mention that 1,4-DCB was considered a category 2 gas until much later in that document (when the inhalation carcinogenicity is discussed), so the reader of the inhalation RfC section of the proposed IRIS entry may have difficulty understanding why the HEC is derived one way for mice and another way for rats.

It may be important to consider that the HEC conversion for the olfactory epithelial effect drives the choice of critical endpoint. If not for the factor of 0.16, the mouse testes mineralization endpoint would yield the lowest BMD/BMDL and the reasons for discounting that estimate (e.g., that it is well below the experimental concentrations) would need to be carefully reconsidered. In fact, I believe some have suggested that for a respiratory effect like the eosinophilic effects in olfactory epithelium seen here, the inhaled concentration itself might be a better basis (or at least as good a basis) for animal-to-human conversion of inhalation exposures. This choice for HEC conversion would make a difference in the selected critical endpoint (even though the numerical estimate of the RfC estimate would not change greatly if the testes mineralization endpoint can be justified on biological/toxicological grounds).

Abigail Jacobs Ph.D.

This was explained OK, but it is not clear how is different anatomy taken into account?

(See Bruce Allen comments on this.)

4. Benchmark dose (BMD) modeling was used to derive the point of departure for determining the RfC. In the absence of information on the level of response to consider adverse, 10% extra risk was used according to the U.S. EPA Benchmark Dose Guidance (U.S. EPA, 2000) as the benchmark response for the dichotomous data sets. Was the correct benchmark response chosen? Was the BMD modeling accurately and transparently described?

Rogene F. Henderson, PhD, DABT, Chair

The choice of the level of response to consider as the BMR is somewhat arbitrary. It is important that the BMR lie within the observable range. Since a 10% extra risk of the eosinophilic changes in the olfactory epithelium was near the low end of the observable range, I consider it an appropriate choice. The BMD modeling was briefly but adequately described and references were given for more detail.

Bruce C. Allen

I think the choice of 10% extra risk is satisfactory in this case. It may be of some interest to note somewhere in the document, that the choice of extra risk (as opposed to additional risk) makes a big difference in the BMD estimates for the endpoints modeled here that have greater than 50% incidence in the control groups. The choice of extra vs. additional risk would make little difference when the background rate is closer to zero. Thus, some specific discussion of the merits of extra risk over additional risk might add to the narrative here.

I have one specific issue with the manner in which the BMD modeling was completed, and that relates to the presentation of all the model results (as in Table B-8) and their comparison/selection. The linear model and the quantal quadratic model are special cases of the Weibull and multistage models. As such, they can be statistically, formally compared to the more general forms. Therefore, comparison of the AICs for the linear and quadratic models to those for all the other models is not appropriate.

The process that I would recommend is that, if desired, the linear and quadratic special cases of the Weibull and multistage models should be compared to the general forms, using the differences in the log-likelihoods and the chi-square-based test of the significance of the "power" parameter of the Weibull model or the degree of the multistage model. If the test of significance suggests that setting the power equal to the special case (1 for the linear model or 2 for the quadratic model) is satisfactory (i.e., there is no significant advantage in allowing that power to differ from 1 or 2 as the case may be), then fix the power (the same comments apply to the multistage degree). If the test of significance suggests that the power should be allowed to vary from the special case (i.e., is not 1 or 2), then allow the model to estimate the power. In either case, the selected Weibull or multistage form should then be included in a tabulation and comparison as shown in Table B-8. The comparison of AICs is an ad hoc method for selecting models that are not "collapsible" (in the sense that the Weibull model, for example, is collapsible to a linear model as a special case). Thus, comparison of models via AIC should not be done when there are formal statistical procedures to determine if one model is preferred over another.

It is clear from examining the GOFP and AIC values in Tables B-8 and B-10 that the best fit for both the Weibull and the multistage models was the linear special case. A formal comparison of the quadratic model to that best fit would reveal that the quadratic model would be rejected. Therefore, the tables should only present rows for the probit, log-probit, logistic, log-logistic,

gamma, Weibull, and multistage models, the "best" model being selected from those based on the AIC.

In the cases of both the eosinophilic changes in rats and the testes mineralization in the mice, the selected models (log-probit and log-logistic, respectively), do appear to be the proper choices despite the fact that the selection procedure ought to be amended as suggested above.

In the case of the eosinophilic changes, I did not see a rationale or discussion for selecting to model only the combination of moderate or severe effects. Almost all the animals (including controls) had some level of eosinophilic effect. Why was the decision made to model moderate or severe as opposed to severe only? And, given the graded nature of the responses, would it not have been better to use a model (like a categorical regression model) that could represent all of the severity levels and their changes as a function of dose? Then the decision about the grade of response could be based on a model that characterizes the dose-related changes in all the severity levels. At the very least, I would expect to see some discussion of the reasons a graded-response analysis was not done or perhaps not even considered.

Abigail Jacobs Ph.D.

(See answer to Question 4 by Jacobs.)

5. Have the uncertainties in the derivation of the 1,4-DCB RfC been adequately characterized? Does the Toxicological Review provide a transparent explanation for the selection of uncertainty factors? Are the uncertainty factors scientifically justified?

Rogene F. Henderson, PhD, DABT, Chair

The usual uncertainties were adequately discussed and the uncertainty factors were well justified. I agree with the choices.

Bruce C. Allen

In general, I think the uncertainty factors have been adequately characterized; the explanation for the specific choices may need to be expanded. Specifically I am concerned that the discussion of the UF for database deficiencies is very brief with little or no characterization of what constitutes a satisfactory "variety of suitable studies." This is especially so when compared to the choice of a UF=3 for this factor for the oral RfD estimate for 1,4-DCB elsewhere in the document. I would like to see specific reasons given for why the RfD required a factor of 3 for database deficiencies whereas the RfC did not require a UF for this feature.

What is entirely unclear to me (but perhaps outside the purview of the requested review) is why having a reported LOAEL in an oral two-generation study of 90 mg/kg-day, as compared to a BMDL of 9 mg/kg-day as the point of departure, should dictate that a UF=3 would be applied (see p. 125 of the Toxicological Review). Again, as in my comments above about using a dose-response basis for picking critical effects, this comparison does not seem to be "apples-to-apples" since in one case the estimate is from a BMD analysis whereas in the other it is apparently a NOAEL/LOAEL determination.

Moreover, I am not at all certain that I understand why inhalation studies would not be considered when evaluating the deficiencies of the data base, even when an oral RfD is being determined. Can one not make any inferences about the information provided by oral studies for the inhalation

route or vice versa? Such inferences were made in other places (in considering the MOA data for the cancer assessment).

Abigail Jacobs Ph.D.

4. Benchmark dose and 5. Uncertainty factors:

They are accurately and transparently described:

See above comments on whether the endpoints selected are meaningful. I have a problem with such a high background rate and the effect does not appear to be de novo, but rather an enhancement of a background rate, and the background lesion is not a background in humans.

(See also Bruce Allen comments on this.)

Additional Comments

Bruce C. Allen

I have a question about the summarization of the "critical effect" levels for 1,4-DCB shown in Table 4-12. Some of the entries present LOAELs with no NOAELs, or vice versa. When there is an identified LOAEL at a dose other than the lowest non-zero dose, does that not imply that lower doses are NOAELs? Similarly, when there is an identified NOAEL at a dose less than the highest dose, does that not imply that (some) higher doses are LOAELs? The presentation of the NOAELs and LOAELs does not appear to be consistent because of the lack of consistency with respect to those designations.

Inhalation Cancer Assessment for 1,4-DCB

1. The principal study used as the basis for the quantitative inhalation cancer assessment is the JBRC 2-year inhalation bioassay (Aiso et al., 2005). Is this study the most appropriate selection for the principal study (i.e., the best study upon which to determine the point of departure)? Has the rationale for this choice been transparently and objectively described?

Rogene F. Henderson, PhD, DABT, Chair

I think this was an appropriate choice. The study was done by inhalation following a standard protocol that is used for cancer bioassays. GLP was followed.

Bruce C. Allen

As in the case of the inhalation RfC derivation, the JBRC is the only full-lifetime inhalation study and is therefore the appropriate selection. Section 5.3.3.2.1 of the Toxicological Review transparently describes this rationale.

Abigail Jacobs Ph.D.

The best (and only) inhalation study is the JBRC inhalation study. The rationale for the choice is properly described. Oral NTP studies can also inform on systemic effects and should not be ignored.

2. An inhalation unit risk was derived using BMD modeling to define the point of departure followed by linear low-dose extrapolation below the point of departure.

(a) Has support for the use of a linear low-dose extrapolation been objectively and transparently presented?

Rogene F. Henderson, PhD, DABT, Chair

No, I do not think so. A linear model is generally used when one thinks the compound is genotoxic and that there is no dose that is "safe." There are many things that suggest that 1,4-dichlorobenzene is not genotoxic. These data are summarized on page 19 of the IRIS document. (II. A. 4). Also the data from the oral study by NTP and from the inhalation study by Aiso indicate that the tumor responses were at the high dose level only. As stated on page 19 of the IRIS document, "The minimal evidence for genotoxicity of 1,4-dichlorobenzene is consistent with the IARC (1999) conclusion that there is weak evidence for the genotoxicity of 1,4-dichlorobenzene in mammalian cells in vitro, and that no conclusion can be drawn from the in vivo data." On the other hand, there is strong evidence for a mechanism involving stimulation of hepatic cell proliferation. The increase in tumors parallels the increase in organ weigh, in support of the cell proliferation MOA. There is a sentence on page 110 of the Tox Review that states, "available evidence indicates that the mechanism leading to the formation of mouse liver tumors following 1,4-dichlorobenzene exposure is based on sustained mitogenic stimulation and proliferation of hepatocytes. Some of the data indicate that the cell proliferation may be a threshold response to cytotoxicity, which would be consistent with the results of the NTP (1987) bioassay." Whatever the MOA for the cell proliferation, whether through cytotoxicity or some other molecular mechanism, it only occurred at the high level exposure in the NTP study. I find the final sentence on page 110 to be a rather weak justification for using a linear model. The evidence strongly suggests a non-linear model. At the very least both a linear and non-linear model should be described.

Bruce C. Allen

The description in Section 5.3.3.2.2 of the choice of data sets is a bit less than transparent. The rationale for selecting to model hepatocellular adenomas and carcinomas together is clearly described at the beginning of the first paragraph on p. 138. But then the discussion describes a significant trend for carcinomas alone in male mice, suggesting that carcinomas alone should be modeled but for males only, apparently; the females also showed a significant trend for carcinomas considered by themselves, yet that endpoint was not modeled separately in females. The document suggests that in an oral study (NTP, 1987), male mice had not demonstrated an increase in carcinomas alone and so adenomas and carcinomas would be modeled together as well. But had not that decision already been made (because of the usual and general reasons for combining carcinomas and adenomas presented at the top of that paragraph)? Also, does it not appear to be a bit inconsistent to cite an oral study here, when elsewhere in the document there

has been a concerted (and in my opinion, over-done) effort to separate oral and inhalation studies?

In addition, the decision to model combined bronchoalveolar carcinomas and adenomas in females is a bit too succinct in its rationale to combine them "consistent with the discussion above" (p. 139, line 2). The "discussion above" is full of conditions and in fact did not end up with only combined carcinomas and adenomas, at least not in the case of the male mice. So how can the decision to only model combined bronchoalveolar carcinomas and adenomas be consistent with that?

I would recommend selection criteria that were more consistent across sex and tumor site. If it is generally considered proper to model adenomas and carcinomas together because the distinction between them is often arbitrary and is difficult to consistently apply, then do not separate them out in some cases but not others. If the motivating factor is to include all cancers with a significant dose-response (including toxicologically supported combinations such as adenomas and carcinomas), then say that and apply it consistently. Neither rationale was applied consistently here.

With respect to the decision to use a linear low-dose extrapolation method: this is the conservative approach, as compared to a nonlinear/UF approach. The review states that 1,4-DCB may not be DNA-reactive (the majority of the genotoxicity tests appear to support this conclusion), but great weight appears to be given to a small subset of the data base that suggests that a purely nongenotoxic/mitogenic effect may be questionable (because of lack of continuing effect with long-term exposure, species inconsistencies). The public review comments were very consistent in the other direction: supporting the preponderance of studies showing lack of genotoxicity and showing consistency between doses producing cell proliferation and those showing increased cancer incidence. Thus, at the very least, the toxicological review and IRIS documents need to have a much more focused discussion of the weight of evidence, including specific reasons for giving greater weight to some portions of the data base (e.g., those subsets not suggestive of a mitogenic effect). A mere listing of the studies and their findings is not sufficient; a critical review and comparison is needed.

However, even if a linear low-dose approach is adopted, the calculation of unit risks based on summing unit risks from two separate tumor types is ill-advised. In the first place, since the entire exercise makes no practical difference (i.e., does not give a final unit risk estimate different from what would be obtained by looking at the "most sensitive" tumor type alone), why bother to present a method that is ad-hoc at best?

The difficulties with the ad-hoc method for combining the unit risks are based on the following observations. It is assumed that the estimates of the unit risk are normally distributed around the maximum likelihood estimate with, for example, the 95% UCL for the unit risk being equal to the MLE (mean) plus 1.645 times the standard error. This is wrong-headed for several reasons. First, in the estimation (model-fitting) procedure it is the BMC that is estimated and the BMCL is the 95% LCL for the BMC. That does not imply that the unit risk (defined as 0.1/BMCL) is the 95% UCL for the ratio 0.1/BMC. Moreover, as seen by the definition of the unit risk, it is a ratio of a fixed value divided by an estimate of some parameter (the concentration associated with 0.1 extra risk); there is no reason to think that such a ratio would be normally distributed around a mean value even if BMC estimates were normally distributed around a mean (MLE). In fact, considering that the ratio is constrained to be positive (the BMC must be positive by definition), it is clear that the assumption of normality is inappropriate. The estimation of the BMCL in software such as BMDS does *not* make such simplistic assumptions (e.g., that the BMC estimates

are normally distributed about an MLE); it uses a profile likelihood procedure that identifies the likelihood of various BMC values and selects the smallest value that gives a likelihood that could not be rejected with 95% confidence.

That last observation in fact suggests the appropriate way to combine tumor types to get a "combined" unit risk. It is based on simulation. For each iteration of the simulation the following would be done. For each tumor separately, one should sample a value from the distribution of BMC estimates consistent with the data in question. That sampling would be based on the corresponding likelihood associated with the possible BMC values, and so it would be consistent with the profile likelihood approach used to get BMCL estimates. Then the separate unit risks would be calculated from the sampled BMC and added together. That sum would represent one possible value of the combined unit risk. If that process is repeated many times (many such iterations) a distribution of combined unit risks would be generated and the 95th percentile of that distribution could be chosen to represent an upper bound for the combined unit risk.

Therefore, the contention starting the third paragraph of p. 141 – that a "statistically appropriate approach" has been used – is wrong. While it is appropriate to add MLE estimates to get a combined MLE, the footnote supporting the contention that a "statistically appropriate approach" has been used concerns the calculation of an upper bound that is wrong for all the reasons presented above. The implication that this is statistically reasonable needs to be removed from the document.

Which gets back to my first point about all this: since the combination makes no practical difference and the simulation approach would not necessarily be straight-forward (it could not be done with BMDS), why bother with it in the first place. I think that it would be sufficient to just point out that the "most sensitive" tumor type gives a BMCL and unit risk that differ by about an order of magnitude (or more) from those that were estimated from the other tumor type under consideration. A quantitative calculation of the combined MLE estimates, *only*, would suffice to show the lack of impact on the final estimate.

There is one other problem related to choice of data for the analyses that are presented, for some of the modeled endpoints. Looking at Appendix B, section 7, and specifically at the BMDS outputs presented there, it appears that the male histiocytic sarcoma data and the female hepatocellular adenoma and carcinoma data were modeled using only 3 dose groups (see pages B-34 to B-37). It appears that the two lowest positive dose groups were combined (using total incidence for those groups and an average concentration value). Why was this done? And more importantly, why was this not discussed anywhere in the document? The only reference to an average dose is buried on p. B-27 (second paragraph), and that only presents an average for the female mouse lung tumor HECs, which was not used in the analysis (pp. B-38 to B-39). This is a major issue that needs to be addressed and corrected for the inhalation risk assessment.

<u>Abigail Jacobs Ph.D.</u>

Re: A linear low-dose extrapolation. The rationale for a linear low-dose extrapolation was not clear in view of the body of data suggesting that 1,4-dichlorobenzene is not clearly genotoxic and probably mitogenic. This section needs more discussion.

(b) The inhalation unit risk is based on the summed risks of developing liver carcinomas and hepatic histiocytic sarcomas (male mice) or hepatocellular adenomas/carcinomas and bronchoalveolar adenomas/carcinomas (female mice). Have the most appropriate data sets been chosen for derivation of the inhalation unit risk? Has the modeling been accurately and transparently described?

Rogene F. Henderson, PhD, DABT, Chair

I had no problem with this approach. Also, it did not affect the calculated risk more than 10%.

Bruce C. Allen

(See answer to Question (a)).

Abigail Jacobs Ph.D.

I don't understand the scientific basis of summing the risks from the different sites, and relevance to humans may differ markedly. I would look at pooled hepatocellular adenomas and carcinomas, since they are a continuum. I would not pool hepatocellular neoplasms with histiocytic sarcomas, which are not lesions of hepatocytes, but of histiocytes and are a systemic lesion. The pulmonary effect in female mice is very marginal. This section needs more discussion

The fact that the hepatocellular neoplasms were seen in 40% of the control male mice in this study, suggests that these mice have an unusual susceptibility and that quantitative extrapolations will be shaky. The effect does not appear to be a de novo effect of the chemical alone.

For hepatocellular neoplasms, I see an effect dose at 300 ppm and a no-effect dose at 75 ppm in the inhalation study.

3. Inhalation dosimetry methods were used in the calculation of the human equivalent concentration (HEC) based on U.S. EPA (1994). Is the explanation for dosimetry choices in the derivation of the inhalation unit risk scientifically justified and transparently described?

Rogene F. Henderson, PhD, DABT, Chair

I found the description of the dosimetry conversions to be clear and reasonable. I did not understand one sentence. It was the last sentence in the large paragraph in the middle of page 139. What two lower exposures are they talking about?

Bruce C. Allen

As stated in the review of the inhalation RfC, in the absence of recapitulating the entire justification (from EPA, 1994) for treating category 2 gases as category 1 for respiratory effects and category 3 for extra-respiratory effects, there is not much more that can be added to scientifically justify the HEC conversions. That does detract from transparency, perhaps, but interested readers should seek out the referenced methodology. Again, it might be useful to have somewhere in the document a rationalization for why 1,4-DCB should be considered a category 2 gas (i.e., its properties with respect to reactivity and solubility that confirm such a categorization, as opposed to just noting that effects were observed both in the respiratory tracts and at remote sites).

The proposed IRIS entry is not consistent with the Toxicological Review in this regard. Although the female mice bronchoalveolar tumors are mentioned in the text (and the HEC conversion for those tumors is described), no data showing the incidence for those tumors and the corresponding HEC values is provided.

Abigail Jacobs Ph.D.

I follow the dosimetry choices.

4. Have the sources of uncertainty been adequately and transparently described?

Rogene F. Henderson, PhD, DABT, Chair

I thought the uncertainties were clearly discussed.

Bruce C. Allen

The discussion of sources of uncertainty is very brief. The discussion of the model uncertainties centers, appropriately, on questions about MOA, but it says little more than that there is some uncertainty about whether or not a linear low-dose extrapolation is reasonable. There is no discussion about how that uncertainty might be resolved or how the decision whether or not to apply linear extrapolation could be improved. Given the above comments about the need for a better weight-of-evidence discussion concerning MOA, the uncertainties associated with determining a MOA could be folded into that discussion.

An uncertainty that is not really discussed is the uncertainty associated with having only a single study adequate for lifetime cancer risk estimation. Even though the study included two species, the lack of any additional information results in uncertainties, but perhaps those are so common in cancer risk assessments that they are not worth mentioning.

On the other hand, the discussion of parameter uncertainties is definitely lacking. Although it is correct to point out that such uncertainty can be visualized by looking at the relationship between BMC estimates and their lower bounds, the statement that "the data were more uncertain than the model fits suggest" (p. 142 last paragraph) is confusing. Uncertain *data* would arise if there were issues about whether the right measurements had been recorded or perhaps if the quality of data collection or experimental design was suspect. That has nothing to do with models fit to those data. Barring such recording, quality, or design concerns, the data are just what they are and are not subject to uncertainty concerns. Assuming binomial observations (as is done for dichotomous responses), the *variability* in the data is a simple function of the proportion responding – the closer those proportions are to 0.5, the more variable will be the observed proportion responding. Again that is not related to any model fits. On the other hand, models that are fit to those data, even if assumed to be the correct models, may not be estimating the "right" parameter values because the variability in the data affects point estimates of those parameters. That is what confidence limits are meant to address.

Therefore, I can not understand the relevance of evaluations of the differences in BMCs and BMCLs obtained when different subsets of the data are fit. As mentioned above, fits of models to the data have nothing to do with data uncertainty or with data variability. Moreover, it is hardly surprising that fitting models to what are essentially different data sets would change the estimates, especially when omitting an entire dose group excludes roughly a quarter of the

observations and *all* the observations at certain points along the dose-response curve. There are valid data-deletion approaches to characterizing parameter estimation uncertainty, but they are not as heavy handed or arbitrary as those described in this document. The discussion currently in the document adds nothing to the evaluation of parameter uncertainties.

As an aside, the concluding statement, "Consequently the recommended unit risk includes all of the exposure groups," does not seem to be entirely consistent with the fact, noted above, that some dose groups were combined when fitting the multistage model. In a sense all groups were included, but in a way that is not standard (averaging of doses in two groups and the combination of their response rates) and in a way that is not explained or documented.

So, the entire discussion of sources of uncertainty seems to boil down to two simple statements – there is uncertainty in the dose-response model and extrapolation method chosen, and there is uncertainty in parameter estimation for the one model chosen. These do not really address the *sources* of those uncertainties, which would be primarily related to the lack of sufficient information to make decisions about MOA.

Abigail Jacobs Ph.D.

Uncertainty in general has been described, but uncertainties relating to this particular compound could be described. (See Bruce Allen comments.)

5. It is EPA's judgment that there is insufficient evidence to establish a mode of action for mouse liver tumors, and thus a linear low-dose extrapolation model was used for the quantitative dose-response assessment. There is evidence, however, that suggests that sustained mitogenic stimulation and proliferation of hepatocytes may be involved in the induction of mouse liver tumors, and that this cell proliferation may be a threshold response. Based on what is known about the mode of action, does the science support EPA developing a nonlinear dose-response model as well to help characterize the cancer dose-response? If so, please provide us with advice on conducting such a nonlinear analysis.

Rogene F. Henderson, PhD, DABT, Chair

I strongly agree that EPA should develop a non-linear model for comparison with the linear model. I think there are enough data to suggest a nonlinear mechanism that a nonlinear model should be considered. If one plots the oral and inhalation absorbed dose against the tumor incidence, one can find a NOAEL or LOAEL and divide by appropriate uncertainty factors.

Bruce C. Allen

My comments above suggest that there needs to be a fuller and more transparent discussion of the entirety of the data base related to mode of action, and that the reasons for EPA's decision to dismiss a large fraction of that data base in favor of a subset that raises issues with a nongenotoxic mechanism need to be explicitly listed and justified. It is on that basis that a decision to do, or not do, a linear extrapolation needs to be made.

Once that decision is made, the appropriate extrapolation method should be applied. I would not be in favor of presenting a linear and a nonlinear extrapolation and comparing the two. Pick the one approach that is consistent with the toxicological judgments, and use it.

If a nonlinear approach were to be chosen, the best method would be to apply uncertainty factors to a point of departure derived from modeling the selected endpoints. The point of departure should be in or near the observed dose range, although it being somewhat below that range should not be a tremendous concern. Selection of a BMD/BMDL for a response of 10% extra risk should be adequate.

I recommend the point of departure approach over using a model to directly extrapolate to low doses/low risks. Unless a much more detailed biologically based model were developed, the confidence in the model predictions for low doses/low risks would be very low itself. Given the generally acceptable criteria that have been applied to develop uncertainty factors for the RfC derivation, those same criteria could probably provide a reasonable basis for uncertainty factors that could be applied in a nonlinear extrapolation.

Abigail Jacobs Ph.D.

(See answer to Question 4 by Jacobs.)

The science supports a nongenotoxic mode of action and a nonlinear low dose analysis. There may be a nonlinear response as well for toxicity secondary to covalent binding of epoxide to proteins. Depletion of glutathione has a threshold in humans (p. 17); there seem to be species/strain differences in metabolism. There can also be differences in epoxidase levels and glutathione levels between mice and humans.