⊕EPA

AN SAB REPORT: DEVELOPMENT OF THE ACUTE REFERENCE EXPOSURE

REVIEW OF THE DRAFT DOCUMENT *METHODS FOR EXPOSURE-RESPONSE ANALYSIS FOR ACUTE INHALATION EXPOSURE TO CHEMICALS: DEVELOPMENT OF THE ACUTE REFERENCE EXPOSURE* (EPA/600/R-68/051) BY THE ENVIRONMENTAL HEALTH COMMITTEE OF THE SCIENCE ADVISORY BOARD (SAB)

November 23, 1998

EPA-SAB-EHC-99-005

Honorable Carol M. Browner
Administrator
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460

>   Subject: Review of the draft document *Methods for Exposure-Response
>   Analysis for Acute Inhalation Exposure to Chemicals:
>   Development of the Acute Reference Exposure* (EPA/600/R-
>   98/051, April 1998)

Dear Ms. Browner:

        In an effort to provide toxicity values for acute noncancer risk assessment for inhalation
exposures, the U.S. EPA National Center for Environmental Assessment has developed a
methodology for performing dose-response assessments for nonconcur effects due to acute
inhalation exposures. The methodology describes the derivation of Acute Reference Exposures
(ARE), a chemical-specific acute exposure (with an uncertainty spanning an order of magnitude)
that is not likely to cause adverse effects in a human population. These estimates, applicable to
single continuous exposures for up to 24 hours, are expected to have wide applicability in
assessing potential health risks due to short-term exposures to airborne chemicals in the
environment. As they are developed and reviewed, the EPA plans to make Ares available to the
public in chemical-specific files found in U.S. EPA's Integrated Risk Information System (IRIS)
database.

        At the request of the Office of Research and Development (ORD), the Environmental
Health Committee (EHC) of the Environmental Protection Agency's Science Advisory Board
(SAB) reviewed the subject document in public meeting, on June 10, 1998 at the EPA's
Environmental Research Center in Research Triangle Park, North Carolina. The draft document
had undergone both internal and external peer review and was revised prior to the SAB review.

        The Office of Research and Development requested that the EHC provide comment on
each of the following aspects of the acute reference exposure methodology:

>   a)   The ARE methodology recommends three approaches for deriving ARE values
>   and describes the types and amount of data that should be used to support each
>   approach. Are these approaches appropriate for deriving acute exposure values?
>   Are the recommendations for types and amount of data appropriate?

b)   The ARE methodology recommends using dosimetric adjustments to derive human equivalent concentrations from animal exposures. The ARE methodology departs from the RfC methodology by recommending default dosimetric adjustment factors of 1.0 for all categories of gases. For particulates, the same adjustments used for developing RfCs are recommended. Does the documentation provide sufficient rationale for these recommendations? If not, please comment on the elements that are lacking. Are the recommended dosimetric adjustments applicable to acute exposure scenarios? If not, please recommend dosimetric adjustments that are more applicable to acute exposures.

c)   The categorical regression option of the ARE methodology involves assigning ordinal severity categories to effect data from toxicity studies that use a variety of species, exposure concentrations and exposure durations. Regression analysis is then used to relate the severity of response to exposure concentration and duration for the entire array of data. For determining the severity category of acute health effects, the ARE methodology document recommends using toxicological judgment rather than a well-defined scheme because such schemes are unlikely to be applicable to a variety of toxic endpoints. Is the expert system for categorizing severity sufficient? If not, how can it be improved?

d)   The ARE methodology recommends using severe effect data, including lethality, for the categorical regression approach, but advises against using lethality and other nonsensitive endpoints when using No-Observed-Adverse-Effect Level (NOAEL) and benchmark concentration approaches. The categorical regression model uses severe effect data to determine the slopes of the probability curves for each severity, the intercepts for the curves and the distance between the various severity curves. Is the guidance offered for including lethal and severe effect data for ARE derivation sufficient? Can the panel suggest ways in which severe effect data could be better utilized?

e)   CatReg software allows individual data and data reported as group information to be combined in a single analysis. The *CatReg Software User Manual* offers three alternatives for placing group and individual data on "equal footing": the use of a scaling factor, g, to accommodate within-group correlations and group size; converting individual data to group data; and estimating individual responses from group information. No alternative is described as preferred. Does the panel have an opinion as to which alternative may be preferable?

f)   In categorical regression, the rules of probability constrain the probability curves for the various severities to be parallel. Although parallelism is a mathematical constraint, it implies the biological interpretation that similar mechanisms of action and kinetics are active in all severity categories. Does the panel view this as a

limitation to the categorical regression approach? If so, how should the use of categorical regression be constrained?

g)   Of the approaches recommended for ARE derivation, categorical regression is the only approach for which duration extrapolation is not required. The NOAEL and benchmark dose/concentration methods (BMC) approaches can only be applied to exposure durations for which data are available. AREs for other exposure durations must be derived by duration extrapolations. AREs for other exposure durations must be derived by duration extrapolations. For extrapolation from short duration values to longer durations, a concentration multiplied by time adjustment is recommended. For extrapolations from long durations to shorter durations, use of the same concentration identified at the longer duration is recommended. These are conservative duration adjustments. Are these duration adjustments appropriate for the approaches to which they are applied? Can the panel suggest other adjustments that may be more appropriate?

The Environmental Health Committee commends the Agency for developing methodology to derive Acute Reference Exposures (ARE). Overall, the NOAEL and the benchmark concentration approaches were found to be well described. In addition, the Committee found the approach for duration extrapolation to be clear and appropriate. The statistical software package seems to be versatile and well put-together.

The EHC had major concerns with the Categorical Regression (CatReg) approach for developing an ARE and, therefore, does not support the use of CatReg in deriving acute reference exposure values. The EHC's concerns with the CatReg approach are based on the lack of biological plausibility for the methodology, the lack of justification for the scaling factor to accommodate within-group correlations, and group size, and the unreliability of the types of confidence limits used. Also, the Agency did not present an example where the dose-response line produced by the categorical regression model adequately described the biological data. Regarding the statistical methodology underlying the CatReg approach, the EHC recommends that the Agency also validate its assumption that all probability curves for the various severities are parallel. During the public meeting, the EHC recommended that the Agency reassess the database to determine the applicability of categorical regression, the basis for this determination with, if appropriate, examples of its usefulness with specific chemicals, and then return to the SAB for a follow-up review of a revised ARE methodology.

The EHC found the expert system for categorizing severity to be inadequate due to the reliance on only a few toxicologists to make decisions on severity of both animal and clinical responses. There was a difference in opinion amongst the EHC regarding the usefulness of lethal and severe effect data in deriving the ARE using the NOAEL and BMC approaches. Some of the committee members supported the inclusion of lethal and severe effect data while others did not. The EHC also found that the BMC and NOAEL/LOAEL calculations did not deal adequately with risk to children.

The EHC was concerned about the Agency's use of expert judgment to categorize severity of effects. It is very difficult to ensure that severity grades across all the different possible endpoints are appropriately and consistently ranked. The Agency should consider holding a workshop to discuss the scientific merit of guidelines for defining severity categories. Since expert judgment systems are usually a function of the experts chosen, guidelines for defining severity categories should minimize, to the extent possible, any degree of subjectivity.

The Committee appreciates the opportunity to review the methodology for deriving acute exposure reference exposures and looks forward to receiving a written response from the Director of the Office of Research and Development's National Center for Environmental Assessment.

Sincerely,

/signed/
Dr. Joan M. Daisey, Chair
Science Advisory Board

/signed/
Dr. Emil A. Pfitzer, Chair
Environmental Health Committee
Science Advisory Board

/signed/
Mark J. Utell, Co-Chair,
Environmental Health Committee
Science Advisory Board

NOTICE

This report has been written as part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

ABSTRACT

The Environmental Health Committee (EHC) reviewed the EPA's methodology document, *Methods for Exposure-Response Analysis for Acute Inhalation Exposure to Chemicals, Development of the Acute Reference Exposure*. The EHC commends the Agency for developing methodology to derive Acute Reference Exposures (ARE), a chemical-specific acute exposure (with an uncertainty spanning an order of magnitude) that is not likely to cause adverse effects in a human population. Overall, approaches for the NOAEL and the benchmark concentration and for duration extrapolation were found to be clear and appropriate. However, the EHC does not support the use of the categorical regression (CatReg) approach for developing an ARE based on the lack of biological-plausibility for the methodology, the lack of justification for the scaling factor to accommodate within-group correlations and group size, and the unreliability of the types of confidence limits used. Also, the Agency did not determine the applicability of categorical regression or provide a basis for this determination with examples of its usefulness with specific chemicals. Regarding the statistical methodology, the EHC recommends that the Agency validate its assumption that all probability curves for the various severities are parallel. The EHC found the expert system for categorizing severity to be inadequate due to the reliance on only a few toxicologists to make decisions on severity of both animal and clinical responses. A workshop to discuss the scientific merit of guidelines for defining severity categories was recommended. The EHC also found that the calculations are lacking in defining risk to children. At the end of the meeting, the Committee recommended that the Agency reassess the database to determine the applicability of categorical regression, the basis for this determination with, if appropriate, examples of its usefulness with specific chemicals, and then return to the SAB for a follow-up review of a revised ARE methodology.

Keywords: acute reference exposure (ARE) methodology, benchmark dose, categorical regression (CatReg) , no-observed-adverse effect (NOAEL)

Dr. Richard Schlesinger, Professor, Department of Environmental Medicine, New York University School of Medicine, Tuxedo, NY

Dr. Ron Wyzga, Technical Executive, Air Quality Health and Risk, Electric Power Research Institute, Palo Alto, CA[1]

Science Advisory Board Staff
Ms. Roslyn A. Edson, Designated Federal Officer, U. S. Environmental Protection Agency, Science Advisory Board (1400), 401 M Street, SW, Washington, DC 20460

Mr. Samuel Rondberg, Designated Federal Officer**,** U. S. Environmental Protection Agency, Science Advisory Board (1400), 401 M Street, SW, Washington, DC 20460[2]

Ms. Wanda Fields, Management Assistant, Environmental Protection Agency, Science Advisory Board (1400), 401 M Street, SW, Washington, DC 20460

---

[1] Unable to attend the meeting but participated in the report writing effort.
[2] Did not attend the public meeting but provided editorial support for this report.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

The EPA developed the methodology to derive Acute Reference Exposures (ARE), a chemical-specific acute exposure (with an uncertainty spanning an order of magnitude) that is not likely to cause adverse effects in a human population. The AREs are to provide toxicity values for acute noncancer risk assessments for inhalation exposures. This methodology was developed by the U.S. EPA National Center for Environmental Assessment (NCEA). The Agency plans to make the AREs available to the public in chemical-specific files that are found in the U.S. EPA's Integrated Risk Information System (IRIS) database.

The AREs are applicable to single continuous exposures up to 24 hours. It is anticipated that these estimates will have wide applicability in assessing potential health risk due to short-term exposures to airborne chemicals in the environment. In the Agency's methodology document, *Methods for Exposure-Response Analysis for Acute Inhalation Exposure to Chemicals, Development of Acute Reference Exposure* and in the supplementary documents, *CatReg Software Documentation* and *CatReg Software User Manual*, the EPA describes three methods for deriving AREs: no-observed-adverse effect level (NOAEL), benchmark concentration, and categorical regression (EPA, 1998a,b,c).

The Office of Research and Development requested that the EHC provide comment on each of the following aspects of the acute reference exposure methodology:

a)   The ARE methodology recommends three approaches for deriving ARE values and describes the types and amount of data that should be used to support each approach. Are these approaches appropriate for deriving acute exposure values? Are the recommendations for types and amount of data appropriate?

b)   The ARE methodology recommends using dosimetric adjustments to derive human equivalent concentrations from animal exposures. The ARE methodology departs from the RfC methodology by recommending default dosimetric adjustment factors of one for all categories of gases. For particulates, the same adjustments used for developing RfCs are recommended. Does the documentation provide sufficient rationale for these recommendations? If not, please comment on the elements that are lacking. Are the recommended dosimetric adjustments applicable to acute exposure scenarios? If not, please recommend dosimetric adjustments that are more applicable to acute exposures.

c)   The categorical regression option of the ARE methodology involves assigning ordinal severity categories to effect data from toxicity studies that use a variety of species, exposure concentrations and exposure durations. Regression analysis is then used to relate the severity of response to exposure concentration and duration for the entire array of data. For determining the severity category of acute health

effects, the ARE methodology document recommends using toxicological judgment rather than a well-defined scheme as schemes are unlikely to be applicable to a variety of toxic endpoints. Is the expert system for categorizing severity sufficient? If not, how can it be improved?

d)   The ARE methodology recommends using severe effect data, including lethality, for the categorical regression approach, but advises against using lethality and other nonsensitive endpoints when using No-Observed-Adverse-Effect Level (NOAEL) and benchmark concentration approaches. The categorical regression model uses severe effect data to determine the slopes of the probability curves for each severity, the intercepts for the curves and the distance between the various severity curves. Is the guidance offered for including lethal and severe effect data for ARE derivation sufficient? Can the panel suggest ways in which severe effect data could be better utilized?

e)   CatReg software allows individual data and data reported as group information to be combined in a single analysis. The *CatReg Software User Manual* offers three alternatives for placing group and individual data on "equal footing": the use of a scaling factor, g, to accommodate within-group correlations and group size; converting individual data to group data; and estimating individual responses from group information. No alternative is described as preferred. Does the panel have an opinion as to which alternative may be preferable?

f)   In categorical regression, the rules of probability constrain the probability curves for the various severities to be parallel. Although parallelism is a mathematical constraint, it implies the biological interpretation that similar mechanisms of action and kinetics are active in all severity categories. Does the panel view this as a limitation to the categorical regression approach? If so, how should the use of categorical regression be constrained?

g)   Of the approaches recommended for ARE derivation, categorical regression is the only approach for which duration extrapolation is not required. The NOAEL and benchmark dose/concentration methods (BMC) approaches can only be applied to exposure durations for which data are available. AREs for other exposure durations must be derived by duration extrapolations. AREs for other exposure durations must be derived by duration extrapolations. For extrapolation from short duration values to longer durations, a concentration x time adjustment is recommended. For extrapolations from long durations to shorter durations, use of the same concentration identified at the longer duration is recommended. These are conservative duration adjustments. Are these duration adjustments appropriate for the approaches to which they are applied? Can the panel suggest other adjustments that may be more appropriate?

The Environmental Health Committee commends the Agency for developing methodology to derive the ARE values. Overall, the NOAEL and the benchmark concentration approaches were found to be well described. In addition, the Committee found the approach for duration extrapolation to be clear and appropriate. However, the EHC had major concerns with the categorical regression approach for developing an ARE and, therefore, does not support the use of CatReg in deriving acute exposure values. The EHC is concerned that the CatReg approach is based on the lack of biological-plausibility for the methodology, the lack of justification for the scaling factor to accommodate within-group correlations and group size, and the unreliability of the types of confidence limits used. Also, the Agency did not determine the applicability of categorical regression or provide a basis for this determination with examples of its usefulness with specific chemicals. Regarding the statistical methodology, the EHC recommends that the Agency validate its assumption that all probability curves for the various severities are parallel. The Committee also found the expert system for categorizing severity to be inadequate due to the reliance on only a few toxicologists to make decisions on severity of both animal and clinical responses. There was a difference in opinion amongst the EHC regarding the use of lethal and severe effect data in deriving the ARE using the NOAEL and BMC approaches. Some of the committee members supported the inclusion of lethal and severe effect data while others did not. The EHC also found that the calculations are lacking in much the same way as the BMC and NOAEL/LOAEL calculations in defining risk to children. In this report, the EHC provides a detailed explanation regarding its concerns with the categorical regression methodology and the expert system for categorizing severity. At the end of the meeting, the Committee recommended that the Agency reassess the database to determine the applicability of categorical regression, the basis for this determination with, if appropriate, examples of its usefulness with specific chemicals, and then return to the SAB for a follow-up review of a revised ARE methodology.

2. INTRODUCTION

2.1 Background

The ARE is a chemical-specific acute exposure level estimate for noncancer effects (with an uncertainty spanning an order of magnitude) that is not likely to cause adverse effects in a human population after acute exposure to inhaled chemicals other than criteria air pollutants. For the purpose of the review document (USEPA, 1998a), acute exposure is defined as a continuous or near-continuous exposure for a period of less than 24 hours. The ARE is intended to be used as a tool in health risk assessment for acute exposure to toxic air pollutants as a result of accidental as well as routine releases. In the EPA ARE document, the Agency proposes three methods for deriving acute reference exposures: a) no-observed-adverse effect level (NOAEL); b) benchmark dose/concentration (BMC); and c) categorical regression.

The EPA developed the Acute Reference Exposure (ARE) methodology because it has determined that health assessment of acute inhalation exposures is essential for risk assessment of both routine and accidental releases. The ARE methods were specifically developed to support the needs of assessments required by the Clean Air Act as amended in 1990 (U.S. Congress, 1990). Section 112 of the Clean Air Act Amendments (CAAA) requires the EPA to develop emission standards based on the maximum achievable control technology (MACT) for major sources of 189 listed pollutants. In addition, the CAA requires the evaluations of residual risk from listed source categories after application of MACT, and further regulation, if needed, to provide an "ample margin of safety to protect public health." The CAAA also requires methods for assessing risk of both cancer and noncancer endpoints from acute and chronic exposures caused by routine or accidental releases of toxic chemicals. Consequently, the AREs are also intended for broad use for any inhaled chemicals.

2.2 Charge

The Office of Research and Development requested that the EHC provide comment on each of the following aspects of the acute reference exposure methodology:

a)   The ARE methodology recommends three approaches for deriving ARE values and describes the types and amount of data that should be used to support each approach. Are these approaches appropriate for deriving acute exposure values? Are the recommendations for types and amount of data appropriate?

b)   The ARE methodology recommends using dosimetric adjustments to derive human equivalent concentrations from animal exposures. The ARE methodology departs from the RfC methodology by recommending default dosimetric adjustment factors of one for all categories of gases. For particulates, the same adjustments used for developing RfCs are recommended. Does the documentation provide sufficient

rationale for these recommendations? If not, please comment on the elements that are lacking. Are the recommended dosimetric adjustments applicable to acute exposure scenarios? If not, please recommend dosimetric adjustments that are more applicable to acute exposures.

c)   The categorical regression option of the ARE methodology involves assigning ordinal severity categories to effect data from toxicity studies that use a variety of species, exposure concentrations and exposure durations. Regression analysis is then used to relate the severity of response to exposure concentration and duration for the entire array of data. For determining the severity category of acute health effects, the ARE methodology document recommends using toxicological judgment rather than a well-defined scheme as schemes are unlikely to be applicable to a variety of toxic endpoints. Is the expert system for categorizing severity sufficient? If not, how can it be improved?

d)   The ARE methodology recommends using severe effect data, including lethality, for the categorical regression approach, but advises against using lethality and other nonsensitive endpoints when using No-Observed-Adverse-Effect Level (NOAEL) and benchmark concentration approaches. The categorical regression model uses severe effect data to determine the slopes of the probability curves for each severity, the intercepts for the curves and the distance between the various severity curves. Is the guidance offered for including lethal and severe effect data for ARE derivation sufficient? Can the panel suggest ways in which severe effect data could be better utilized?

e)   CatReg software allows individual data and data reported as group information to be combined in a single analysis. The *CatReg Software User Manual* offers three alternatives for placing group and individual data on "equal footing": the use of a scaling factor, g, to accommodate within-group correlations and group size; converting individual data to group data; and estimating individual responses from group information. No alternative is described as preferred. Does the panel have an opinion as to which alternative may be preferable?

f)   In categorical regression, the rules of probability constrain the probability curves for the various severities to be parallel. Although parallelism is a mathematical constraint, it implies the biological interpretation that similar mechanisms of action and kinetics are active in all severity categories. Does the panel view this as a limitation to the categorical regression approach? If so, how should the use of categorical regression be constrained?

g)   Of the approaches recommended for ARE derivation, categorical regression is the only approach for which duration extrapolation is not required. The NOAEL and benchmark dose/concentration methods (BMC) approaches can only be applied to

exposure durations for which data are available. AREs for other exposure durations must be derived by duration extrapolations. For extrapolation from short duration values to longer durations, a concentration x time adjustment is recommended. For extrapolations from long durations to shorter durations, use of the same concentration identified at the longer duration is recommended. These are conservative duration adjustments. Are these duration adjustments appropriate for the approaches to which they are applied? Can the panel suggest other adjustments that may be more appropriate?

3. RESPONSE TO THE CHARGE

3.1 General Comments

The EHC raised several issues regarding terminology and definitions. The specific recommendations are identified in this section of the report.

The Agency should consider modifying the definition of the ARE which appears on page 3, lines 16-18 of the review document. The ARE should be defined as "the *highest* exposure (concentration and duration), with an uncertainty spanning an order of magnitude, that is not likely to cause adverse effects in a human population, including sensitive subgroups, exposed to that scenario on an intermittent basis."

The portion of the document dealing with terms such as adverse effect, threshold, etc. is very confusing. In several places effect level and dose level are used interchangeably. For example, on page six the "threshold effect" is defined as the "minimum adverse effect" in an individual. Does this refer to the effect that occurs at the smallest dose? If so, it should be stated in clearer language. The phrase "in an individual" suggests that the "minimum adverse effect" will be individual-specific rather than chemical-specific. In the same paragraph, directly after a threshold is defined as an "effect", the rest of the paragraph assumes that a threshold is a "dose".

Definitions in the document of terms such as "severity" and "adversity" are essentially non-informative. The document should include a list of adverse effects, or at least some criteria that could be applied to decide when an effect is adverse. Some additional guidance in this area is badly needed. Based on the title, it would be logical for Table 2-1 to contain this type of information. The title of this table is "Definitions of Severity Levels Used in Noncancer Risk Assessment". However, this table actually contains definitions of doses associated with various not-clearly-defined severity levels. Also, severity is identified with incidence (page 23), which is quite confusing.

The document should be modified with more emphasis given to definitions and guidance about classifying effects according to severity. Guidelines should be provided for assigning these severity scores to a small number of severity grades. There could be four or five severity levels (e.g., none, mild, moderate, and severe). In framing these guidelines, EPA needs to address both numerically quantifiable data sets, and non-quantifiable (i.e., qualitative) data sets. Failure to consider such differences, particularly *vis-a-vis* qualitative data, could introduce unwanted variance to severity estimate.

The fraction of individuals having a given severity score would be the incidence of that level of severity. Incidence and severity are different concepts, and intuitively would appear on different axes. The present document leads the reader to confuse these two concepts

The BMC and NOAEL/LOAEL calculations do not differentiate between adults and children. The current risk assessment/risk management process does not adequately take into account exposure/dose differences in children compared with adults, or qualitative differences in sensitive endpoints. It also does not address the fact that for certain exposures and certain endpoints, children may be more resistant than adults. The Agency should reassess the data and then determine if the calculations can be revised to address exposure to children and adolescents. One source worth considering for guidance is the recent report by EPA's Scientific Advisory Panel addressing uncertainty factors (SAP, 1998).

The sensitive endpoints problem is more difficult, as concluded by the authors of the report *Pesticides in the Diets of Infants and Children* (NRC, 1993). There are limited data on inhalation exposure to children, and current toxicological testing does not model for childhood/adolescent exposure. Rarely do studies address sensitive endpoints for fetal exposure.

Although the Committee did not specifically address risk management issues, given such data gaps for pesticides, one of the EHC's members recommends that an additional 10 fold uncertainty factor be applied to the RfCs as well as the AREs; another recommends that the uncertainty factor be included, but that it be allowed to vary (e.g., between 1.5 and 10), depending on the strength of the data for the specific case at hand. On the other hand, other EHC members disagreed that an additional 10-fold uncertainty factor for children be automatically added. A 10-fold uncertainty factor for sensitive population is already routinely added. Adding an additional 10-fold uncertainty factor for children essentially means that a total of a 100-fold uncertainty factor is being added for children. It is essential that, before such a policy decision is recommended, a meaningful scientific discussion of data on relative sensitivities be held.

3.2   Charge 1 - Approaches for developing the ARE and Charge 6 - Parallelism of Severity Curves in Categorical Regression[3]

The acute reference exposure methodology recommends a) the NOAEL, b) the benchmark concentration (BMC), and c) a categorical regression (CatReg) as approaches for deriving acute exposure values. Both the NOAEL and the BMC approaches are familiar processes. The benchmark approach is relatively new and additional development and experience are needed, particularly regarding the application of the benchmark approach to continuous endpoints. However, increasing familiarity with the BMC and its comparison with the NOAEL are serving to provide confidence in both approaches. The relative values of each approach, depending on the nature of the database, are well described. Table 3.1 was found to be particularly helpful in illustrating the data requirements for each analysis..

There was not a consensus within the Committee on the accuracy of the statement made on page 7 of the draft document which states that "... compared to the NOAEL approach, the

---

[3]Because of significant overlap in the relevant issues, these two Charge issues are combined in this section in order promote clarity and avoid repetition.

BMC method has the advantages that it utilizes more information from the concentration-response curve, is less influenced by experimental design (e.g. dose spacing), and is sensitive to the influence of sample size." Some Members suggested that the EPA's discussion of the benefits of the BMC relative to the NOAEL be made more specific and better balanced, incorporating more discussion on the pros and cons of each approach. It was also noted that some toxicologists do consider the entire dose response when using the NOAEL approach. Given the above considerations, both the NOAEL and benchmark procedures appear to be appropriate for developing AREs.

With regard to the BMC methodology, some of the Committee members were of the opinion that: a) the central estimate should be evaluated as it is a more stable and precise estimate of the response at a given dose level especially in the low dose region, b) boundaries should be placed regarding the extent of extrapolation outside the dose response curve, and c) an additional UF for severity should not be added solely to the BMC.

There are major concerns with the categorical regression analysis. The CatReg approach combines data from a variety of species, exposure concentrations and exposure durations. Although the mathematics of the CatReg approach is probably solid, support for the existence of an underlying physiological basis is not clear. The Committee found EPA's assumption that all the probability curves for the various levels of severity are parallel to be particularly troubling. It is premature to implement this method for risk assessment purposes until the assumption on parallelism is validated. Otherwise the analysis will run the risk of being statistically driven with no toxicologic basis.

In addition, statisticians with strong toxicology backgrounds who can address and evaluate the assumptions underlying this approach should review the categorical regression approach with much greater scrutiny. The EHC expressed considerable concern as to how well this approach can be implemented in practice, because it will be extremely labor intensive to determine severity grades for each individual animal or for each group across different endpoints. The validity of this method will depend completely on how well this is done.

The inclusion of all studies indexed by "severity" category introduces a major judgment parameter. What are criteria short of lethality for defining severity? How do human data get categorized in comparison with animal data? It will be extremely difficult to categorize a wide variety of endpoints by a severity analysis and will lead to an "apples and oranges" approach. The draft document's Appendix B illustrates the Committee's concerns with this methodology. A judgment is made that a 30% change in specific airways conductance is a moderate adverse effect and is comparable to what is characterized as a moderate effect of pulmonary edema in animals. Clearly pulmonary edema is a life-threatening effect while this change in lung function is probably of no clinical concern. Such judgment calls lead to inappropriate information fed into the model and an unreliable ARE. The Committee's recommendations on this issue are further elaborated in its response to Charge question 3 on severity categories in Section 3.4 of this report.

Some of the benchmark analyses used a polynomial model, while others used a Weibull model. There is a significant difference of opinion among the Committee regarding the usage of these statistical models. Some members are of the opinion that there is no clear biological justification for one of the statistical models over the other and therefore recommend that the EPA select a default model. Other members of the Committee are of the opinion that the EPA should not choose a default model because doing so would limit scientific information that could be very valuable to the risk assessment process. Those who do not recommend using one statistical model as the default also cite the value of running different models for comparison and interpret the potentially different results as possibly due to a function of model variation as opposed to biological variation.

The BMC is a lower statistical confidence limit on the dose corresponding to a specified level of risk called the benchmark risk (BMR). In applications of the benchmark methodology applied to quantal data, EPA proposes a BMR of 10% to define the benchmark concentration. Although this is primarily a policy/risk management issue, there was a difference of opinion amongst the Committee about defining the BMR at 10% increase as a default value for routine use. Some of the Committee recommend using the 10% at least when the benchmark is based upon animal data, unless there is some compelling reason for deviating. Other members of the Committee were against defining the BMR at 10% increase because they were of the opinion that there is no scientific justification for using 10% for studies or endpoints other than the developmental toxicity studies where 10% was the closest estimate of NOELs. These individuals were also concerned that the selection of an arbitrary value to define the BMR would replace meaningful toxicological evaluation.

In the development of the BMC for cumene, the report (page A-12) defines and uses the term 'extra risk.' However the use of this term is not explained in the report itself and is not discussed in relation to the BMR. It is recommended that this be added to the report.

As noted above, the discussion on the draft document's Pages 54-55 regarding severity and the benchmark approach is confusing. The NOAEL approach also has the same problem of not providing any inherent information about the severity of the effect. It is inappropriate to add an additional uncertainty factor (UF) for severity to the benchmark approach and not for the NOAEL calculations.

The benchmark approach forces isolation of individual effects. In the example of cumene, the approach was applied to "rearing," one of approximately 30 endpoints evaluated in the Functional Observational Battery (FOB). These endpoints should not be evaluated as isolated discrete endpoints. Rather the pattern of effects is what needs to be evaluated. Using the NOAEL approach, the neurotoxicologist would look across all the endpoints to determine if there was a pattern of effect across related endpoints that would be indicative of a treatment related effect and determine the relative severity of the overall symptoms. The benchmark approach forced selection of one endpoint "rearing" which is perhaps one of the less reliable measures in the FOB to evaluate on its own.

10

Finally, there remains concern that the BMC typically identifies a lower or simply different ARE than the NOAEL (much like the RfC analysis). The Agency is encouraged to show and compare both approaches for all ARE analyses.

It is not clear what factors must underlie the assumption of parallel response lines for categorical regression. Consequently, the Committee can not differentiate between biologically driven effects and effects arising from the inherent mathematical structure of the approach. One troubling consequence of the parallelism assumption is that estimated curves for a less adverse effect may be driven by lethality data. The example of hydrogen sulfide in Appendix B provides particular cause for concern. In this analysis there were five studies involving lethal effects and 11 studies involving sub-lethal effects. When the EC-T10s derived for sublethal effects using categorical regression were compared to the benchmark BMC10s obtained directly from the sublethal studies (Table C-5), the EC-T10s were found to be lower than the corresponding BMC10 by huge factors (88 and 197), despite the fact that these quantities should be comparable. Incidentally, the EC-T10, as calculated, is actually a concentration that depends upon exposure duration. The equation (2-5) in the document incorrectly defines the EC-T10 as a time that depends upon concentration. The draft document inappropriately refers to this as exposure (concentration and duration). Furthermore, the notation, "EC-T10," which treats concentration and duration on an equal footing, is confusing and potentially misleading.

It is not clear why results produced by categorical regression are so much more conservative than those derived from the benchmark approach in this example. It may be due in part to the constraints imposed by the assumption of parallel response lines. Another possibility is that the confidence bounds calculated in the categorical regression are Wald-type limits, which are sometimes unreliable. However, this example clearly suggests that categorical regression may not be a reliable alternative to NOAEL or benchmark modeling. In general, it is problematic to combine data from different studies and different protocols and involving different endpoints in a common analysis.

There is not necessarily a net gain by including different categories of responses in a single analysis, as is done in categorical regression. To estimate concentrations corresponding to, say, a 10% increase in a minimal adverse effect, EPA should consider including only the (dichotomous) data on whether or not a subject had that effect. With this approach no parallelism assumption would be required. One potential limitation would be that studies that include information of lethality alone could not be included in the analysis. However, it is not clear that such data contribute meaningfully to estimation of EC-T10's corresponding to less severe effects. EPA generally is not interested in estimating concentrations associated with lethality. However, when such estimates are needed they can be obtained by including only dichotomous lethal responses in the analysis. Although this approach might prove to be more reliable that the current approach of combining data of different severities in a common analysis, it should be recognized that even with this modification, this might still prove to be unreliable because it involves combining diverse types of data in a common analysis, and because of other potential shortcomings discussed below.

The statistical software package seems versatile and well put-together, for the most part. However, confidence intervals in categorical regression are based upon Wald-type statistical theory. These types of confidence limits are often unreliable (Cox and Oakes, 1984). Moreover, they are often strongly dependent upon how the model is parameterized. Confidence limits based upon the asymptotic distribution of the likelihood ratio are generally more reliable, and they are independent of how the model is parameterized (Cox and Oakes, 1984).

If log concentration and log duration are used in a categorical regression model, there is no provision for background response; i.e., the background response is assumed to be zero. Whereas this might often be a reasonable assumption for lethality, less severe responses generally will have a background. The categorical regression approach should account for background response.

If concentration and duration are not transformed in the categorical regression model, the model does predict a background response, but the EC-T10 is for a 10% overall response, uncorrected for background. It should correct for background in the same manner that the benchmark procedure corrects for background. It is possible that the background response of a mild effect could be greater than 10%, in which case the EC-T10 would not be defined.

The document does not present any comparisons of results from using log-transformed or untransformed exposure and duration (In the hydrogen sulfide example, the document does not even report which procedure was used). Since these two options are available in the package, the Agency should compare results from the two approaches.[4] If the differences are significant, EPA should provide guidance on how to handle this issue. Other options (e.g., logit, probit or log-log model) should be similarly investigated. Also, analyses should be developed that compare categorical regression to the NOAEL and the benchmark approaches.

The benchmark and categorical regression approaches are similar in many respects. Categorical regression produces a type of benchmark. Categorical regression is a generalization of benchmark analysis in two ways: it allows concentration and duration to be included separately, and it allows for multiple severity categories. Benchmark analysis treats background responses appropriately, whereas at present categorical regression does not. Although categorical regression is recommended for analyzing data from different experiments simultaneously (meta analysis) and benchmark analysis is not, there is no conceptual reason why the benchmark could not be used in a meta analysis. However, there is some concern whether it is advisable to include all sorts of data -- from different studies with different experimental protocols and from different endpoints -- in a single analysis, as is routinely done with categorical regression.

The relationship between dichotomous data and categorical data should be made clearer. Categorical data are data that can assume only a known finite number of values. Dichotomous data are simply a special case of categorical data in which the number of values (or categories) is

---

[4] Actually there are four possibilities, since one variable could be log-transformed and the other kept untransformed.

two. Although categorical data may be specified in descriptive terms, categorical data are not the same as descriptive data.

The EHC appreciates the intent of the Agency in developing the categorical regression approach, that is, to provide a methodology to determine AREs for specific concentration-time products up to 24 hours without having to rely on conservative defaults regarding applicability of the convention that C x T equals a constant. Unfortunately, the categorical regression is a complex meta-analysis that requires far too many assumptions. The EHC is not convinced that the limited data available can be used in analysis that will be informative and meaningful.

3.3 Charge 2 - Dosimetric Adjustments for Deriving Human Equivalent Concentrations

The Committee is concerned with the methodology underlying the proposed dosimetric adjustment factors (DAF) for both gases and particles. Table 2.5 presents data on pulmonary ventilation and surface area for various species, providing the basis for estimation of an animal to human gas dose ratio for the extrathoracic, tracheobronchial and pulmonary regions of the lungs. This table addresses the assumption that the major interspecies determinants of dose delivered to the respiratory tract are functions of the concentration of the toxicant, the volume of air breathed in, and the total surface area of the respiratory tract. In the examples presented in the draft document, data from this table were used in some cases, but not used in others .

In addition, the use of a default value of 1 for all three gas categories (based upon a different rationale for each category) troubled the Committee. For category 3 gases, the use of a default value of 1 for regional gas dose ratio (RGDR) when partition coefficients for animals and humans are unknown would result in a lower value for human equivalent concentration (HEC) than if higher values of the partition ratios between animals and humans were used. This will usually be a conservative approach. The HEC is for a target organ, that is not the respiratory tract (RT) for this category of gas. For category 1 gases, the default of DAF = 1 is also used. For category 1 gases, the default of DAF = 1 is also used. For the tracheobronchial region of the respiratory tract, this value of one is very close to the animal to human ratios based on lung dynamics as shown in Table 2-4. For the pulmonary region of the respiratory tract, the data in Table 2-4 shower higher ratios for rats and mice, but it was considered prudent (and conservative) to retain the DAF of one. However for the extrathoracic region of the respiratory tract, all of the animal to human ratios in Table 2-4 were markedly less than one, which would lead to a lower LOAEL/HEC. Because experimental data suggested that gas deposition in the extrathoracic region is actually similar in animals and humans, the Agency decided to use the DAF of one (less conservative) unless there is compelling scientific information to do otherwise. It is not clear why the calculated ratios based on a comparison of lung dynamics in animals and humans do not agree with experimental observations for the extrathoracic region of the respiratory tract.

The Committee also found that the Agency seems to ignore gas deposition in the more proximal region when considering the lower respiratory tract. The regional deposition dose ratio (RDDR) for particles uses fractional deposition values to obtain estimated doses in the three

regions of the respiratory tract. This presumably takes into account removal in areas proximal to those of interest, e.g., removal in the extrathoracic (ET) region is accounted for in calculating delivered dose to tracheobronchial (TB). However, in examining the methodology for estimating the regional gas dose (RGD), it is not clear how this regional removal is accounted for when considering those gases which have effects beyond the ET, and for which there are no PB-PK model (which presumably would have such regional estimates) available. Also, in the case of reactive gases which are relatively insoluble in water (e.g., a gas having characteristics similar to ozone), there does not seem to be any dose to certain areas over that which would be estimated solely from ventilation to surface area ratios.

The EHC is particularly concerned that the methodology for dosimetry adjustments in the ARE documents is different than that described in the RfC documents, since both of these methodologies are concerned with estimation of non-cancer risks, and one would expect that a common methodology for dosimetry adjustments and rational for application of default approaches would be used in both documents. We recommend that the authors of both documents develop a common framework for dosimetry adjustments in the two documents.

The statement on clearance kinetics on page 45, lines 26-28 of the review document, may not apply to particles deposited in the upper respiratory tract. And, as noted, clearance kinetics have not been incorporated into the modeling of retained dose for particles. The draft documents states that even if there are differences in clearance rates within the 24 hour time frame (the temporal focus of the ARE methodology) this finding would not be expected to alter estimates of retained dose obtained without consideration of clearance. This is not entirely true. While it may be true for particles which have a major deposition site and major target site in the alveolar region, the dosimetry of particles which deposit largely in the upper respiratory tract and upper tracheobronchial tree (the latter an important site of consideration in some potential critical responses, such as asthma) would certainly be affected in the short term by consideration of clearance kinetics in the modeling effort. Thus, the blanket statement regarding clearance kinetics in the document should be modified.

The discussion above notwithstanding, this section of the review document clearly separated science policy and assumptions from scientific fact. Some additional EHC recommendations on the dosimetric adjustments are:

a)   The Agency should reconsider its statement that gas deposition is the same for animals and humans. The EHC did not concur with that statement.

b)   The Inhalation Reference Concentrations (RfCs) and the Acute Reference Exposures (AREs) for the dosimetric adjustments should be harmonized. The EHC is particularly concerned that the methodology for dosimetry adjustments in the ARE documents is different than that described in the RfC documents. Given that both of these methodologies are concerned with estimation of non-cancer risks, one would expect that a common methodology for dosimetry adjustments

14

and rational for application of default approaches would hold between the two documents. The authors of both documents should develop a common framework for dosimetry adjustments in the two documents.

c)   The statement on page 33 and 34 of the draft document regarding the concentration of the inhaled gases should be rewritten or removed since it is not really an assumption in PBPK modeling.

d)   The Agency should acknowledge in the document that the dosimetric adjustments might not be applicable for highly reactive chemicals, since the EPA only takes into account the partition coefficient.

3.4 Charge 3 - Determination of Severity Category for Health Effects

Based on the examples provided, the Committee found that the expert system for categorizing severity is inadequate. The concept that only a few toxicologists can make decisions on severity of both animal and clinical responses will lead to misjudgments. If the Agency decides to pursue the categorical regression analysis, it would be most appropriate to have a panel of experts including toxicologists, economists, and experienced clinicians involved in the risk assessment process. In addition to the above suggestions, the EHC believes that the determination of severity needs to be on a case-by-case basis, and will evolve through the "case law" process.

An example of where the expert judgment system fails is the hydrogen sulfide (H_2S) case analysis. The Agency should consider holding a meeting or a workshop to discuss the scientific merit of guidelines for defining severity categories. Since expert judgment systems are usually a function of the experts chosen, guidelines for defining severity categories should minimize, to the best extent possible, any degree of subjectivity.

The mathematics underlying the CatReg methodology were not clear to many of the EHC members. In addition, the EHC also had some problems with the gradation of severity between different responses with non-parallel response curves. The model specifications and default were, at times, difficult to comprehend. This approach should be refined so that its reliance on judgment and good science can be maintained and the protocol presented in a manner comprehensible to non-modelers.

The benchmark dose approach aids in estimating a "surrogate" NOAEL when the choice of doses in the study result in the lowest dose being a LOAEL. The benchmark dose approach is also helpful in interpolating a more precise NOAEL that may be closer to the LOAEL than the NOAEL established by the dose spacing chosen for the study. The EHC recommends caution when extrapolating benchmark results beyond the range of experimental doses.

One of the most important differences between RfC and ARE methodology is that the ARE uses both concentration and time as specific factors in defining exposure. Of the three approaches used for ARE, only CatReg actually uses exposure duration for computational purposes. The NOAEL and BMC approaches, however, do require exposure duration (as in X mg/m^3 at Y h)) for proper interpretation and use. Haber's Law accounts for this and it worked for acute exposure to fast acting inhalants (war gases) but it did not fit data from other situations. Druckery (1967) showed that Haber's Law worked for nitrosamines (carcinogens), and Rozman and Storm (1997) have now shown that steady state conditions are required to make Haber's Law fit the broad range of dose and time scenarios. Haber's Law should be predictive when the time factor includes both calendar time and time involved in critical metabolic changes. This idea is alluded to in the document but should be explored and developed since it will have an enormous impact on the ability of studies in the test species to predict results in the target species and improve the ability of toxicology to handle dose, route, pattern, and other exposure parameters.

Expert judgment appears to be the only currently feasible way to categorize severity. Statistical methodologies that could deal with this complex, multi-dimensional issue have not yet been developed. In practice, however, it will be very difficult to ensure that severity grades across all the different possible endpoints are non-subjectively ranked. This problem will likely make the categorical regression approach not feasible for risk assessment. As previously recommended, the Agency should consider holding a meeting or a workshop to discuss the scientific merit of guidelines for defining severity categories. Since expert judgment systems are usually a function of the experts chosen, guidelines for defining severity categories should minimize, to the best extent possible, any degree of subjectivity.

3.5 Charge 4 - Use of Severe Effect Data

There was a difference of opinion whether the lethal and severe effect data should be employed in deriving ARE when using the NOAEL and BMC approaches. Some of the Committee Members supported the inclusion of lethal and severe effect data while others did not. Severe effect data, including lethality, poses difficult problems for risk assessors. On one hand, there is a desire to use the full range of biological/toxicological data, including that related to severe lethal effects/lethality. On the other hand, there is a great deal of uncertainty related to the use of these data. Of special concern is the mechanistic relationship between lethality, for example, and less severe toxic endpoints of more direct relevance for risk assessment. In principle, biologically based dose response models in which the mechanistic response for the continuum from initial biochemical lesion to frank toxicity, including death, are described mathematically, can be used to predict the incidence and severity of effect from mild to severe. The ability to predict these disparate data with one unified model lends credence to the mechanistic underpinnings of the model. In the absence of this kind of modeling approach, it is difficult to justify assumptions that the lethal and mild effects are mechanistically related and with a similar dose response function.

The rationales for supporting the use of lethal and severe effect data include: a) the opportunity to look at all the data to get some appreciation as to how, after severity categorization, the severe effects relate to other effects; and b) the opportunity to conduct a reality check to determine if any ARE should be less than that calculated using severe effects data.

The categorical regression method of deriving ARE values, a method which does make use of severe effect data and which permits the estimation of many different concentration multiplied by time values for AREs, has a good deal of appeal. The process makes use of every bit of data available and also, as stated on page 56, lines 10-12, the underlying premise of the approach is that the severity of the effect, not the specific measurement or outcome incidence, is the information needed for assessing exposure-response relationships for non-cancer endpoints. This approach marks the difference between the approaches used for years in cancer risk assessment, in which incidence is the major, and appropriate focus, and approaches for assessment of non-cancer adverse health effects for which the severity of effect is a major concern.

A positive aspect of the categorical analysis method is that all the available data are graphed on a single chart and one can immediately get a rough picture of the level of the concentration multiplied by time values that can be expected to cause adverse effects of varying severity. The downside of the use of grades of severity in the categorical analysis method is that the probability methods force the assumption of parallel curves for the probability of each of the severity levels. This intuitively does not seem logical. Perhaps more thought could be given to how one might use the data in the graphical presentation. The EPA should focus on the concentration multiplied by time level above which one can expect, with some defined probability, to see an adverse health effect of mild severity. The EPA should also explore whether that could be determined without also determining the probability of the moderate and severe effects. A more detailed discussion of the concentration multiplied by time values is provided in Section 3.2 of this report.

3.6 Charge 5 - Use of Group and Individual Data for Categorical Regression

The use of a scaling factor to accommodate within-group correlations and group size, g, is one of three alternatives described in the *CatReg Software User Manual* for placing group and individual data on "equal footing." The "g" factor is confusing and makes this approach less transparent and probably will not be understood by the majority of risk assessors and toxicologists. In general, the amount of data conversion should be minimized. If the majority of studies provide individual data, the Agency should consider converting the fewer studies with group data into individual data using some approximations based on the mean and standard deviation (S.D.) to plot the data. If the majority of studies are group data, then the EPA should consider converting the individual data into group data.

The EHC was also concerned about the calculation used in the software to derive the confidence limits.

3.7 Charge 6 - Parallelism of Severity Curves in Categorical Regression

Discussion of this Charge issue has been incorporated in Section 3.2 to avoid repetition.

3.8 Charge 7 - Duration extrapolation

The approach used in this section is clear and appropriate. The document must clearly articulate that any studies greater than 24 hours will **not** be used in this assessment. For example, Haber's law will **not** hold if one tries to extrapolate from a 90-day study to a 6 hour exposure.

The problems faced in providing guidance for exposure concentrations (predicted not to cause an adverse (non-cancer) health effect over a specified duration based on data from toxicity tests of a different duration) is quite familiar to several Members of the Committee. For example, some of the EHC Members and Consultants served on the Committee on Toxicology (COT) of the National Research Council. The COT faces this issue frequently. The paragraphs under 2.3.2 of the review document, *Duration Extrapolation*, are well written. The Committee concurs with most of the conclusions, especially the sentence on line 22 of page 47, "It is therefore preferable to determine the duration dependence on a chemical-specific basis." The COT has had a general rule that one can extrapolate forward from a short-term exposure to a longer-term exposure, but one cannot extrapolate backward from a longer-term exposure back to a short-term exposure. This is essentially the procedure described by the EPA. However, there has been reluctance in the COT to extrapolate over too long a time period. Because the EPA is only developing ARE for a 24-hr period, the extrapolations should not be over a long time.

As is correctly stated in the EPA document, Haber's law was developed mainly for lethality as an endpoint, and was originally used to predict the duration required to kill at various concentrations of nerve gas. It was not developed for protecting the public health at low exposure concentrations. Therefore, in each case, it is essential to consider all the mechanistic information available about a chemical and to use that information to predict the dependence or non-dependence of the chemical's toxicity on duration. Structural analogues can provide some information on what one might expect from chemicals for which there is absolutely no data. Standard rules regarding the use of Haber's law should not supplant the necessity for using good judgment based on what is known about the mechanism of action of the material of concern. For example, if a compound is a strong irritant the concentration not the duration of exposure will be of primary importance.

The Agency should include kinetic information in the process when such information is available. The dose of the chemical at the active site will determine the toxicity. Therefore, the "area under the curve" or internal dose is what is of concern. This dose will depend not only upon the amount deposited or absorbed, but also on the clearance rate or rate of activation or inactivation, depending on the mechanism of action. As knowledge of these toxicokinetic parameters becomes known, the information should be incorporated into the ARE process.

Some statements on pages 48-49 are not clear to the Committee. On page 48, lines 5-6, it is stated that a reasonable default is to assume a high value of n to extrapolate to shorter durations. The EHC supports the statement as a conservative approach. However, the next sentence (lines 6-7) does not seem to follow. If a chemical's toxicity were not dependent on time, then the default described in the previous sentence would be exactly right. So why is caution required? The Agency should revise this section of the document.

Besides the severe effects being duration dependent, chemicals that accumulate in the body could be duration dependent in their toxicity. The EPA should include this issue in the discussion.

On page 49, lines 17-19, one must consider not only irritants, but also chemicals that can have an anesthetic effect at high concentrations, as possibly causing decreased ventilation rates.

3.9 Benchmark Dose Approach

The Charges to the SAB did not include any specific reviews of the benchmark dose approach. Since it may be premature to apply the categorical regression for regulatory purposes in the near future, it is likely that the NOAEL or BMD approach will be more frequently used. Thus, the EHC highly recommends that the Agency address the following concerns/questions regarding the benchmark dose approach.

On page 7 of the review document, the BMC is defined as the lower bound on the concentration predicted by the model to cause a defined response. For pedagogical reasons, the EHC recommends that the benchmark dose be defined simply as the dose corresponding to a specified increase in risk. This definition is independent of the method used to estimate the benchmark dose, and, in particular, whether an MLE or a statistical lower bound is used. The EHC also recommends that the Agency present both point estimates and statistical confidence bounds for the benchmark dose to allow the risk manager to assess the degree of uncertainty associated with the benchmark estimate.

Page 7 states that "compared to NOAEL approach, the BMC method has the advantages that it utilizes more information from the concentration-response curve, is less influenced by experimental design (e.g., dose spacing), and is sensitive to the influence of sample size." There was a difference of opinion amongst the Committee regarding the accuracy of this statement. Some of the Committee recommends that this discussion on the benefits of the BMC relative to the NOAEL be made more specific and more balanced.

Page 55 states, "In general, the BMC should be treated as equivalent to a NOAEL in the derivation of an RfC, but that an additional UF for severity may be considered, particularly for continuous data." This statement should be removed as there is no scientifically justifiable reason to add an additional UF to BMC and not to the NOAEL.

The EHC recommends that the Agency place boundaries on the extent of extrapolation outside the dose response curve. The document notes that the choice between BMC, NOAEL and CatReg approaches should be matched to and dictated by the data available for a particular chemical so as to optimize the use of the data. However, in describing the pros and cons of each approach, the Agency should properly acknowledge that a significant amount of scientific judgment and evaluation of different effects across entire dose response curves is integrated into the assessment of the NOAEL approach.

It should be recognized that the BMD approach as outlined by the EPA would result in a more conservative risk assessment than the NOAEL approach. The paper cited on page 53 by Allen *et al.*, (1994) indicates that on average, the NOAELs are three times greater than the LED_{10} and 6 times greater than the LED_{05}. The Agency proposed to add an additional UF for severity when just the BMD approach is used. If we assume a 3 to 10 fold UF for severity, then the BMD approach will result, in effect, in a dose level that is generally 10 to 30 fold lower than if the NOAEL was used for risk assessment.

3.10 CatReg Software

On page 12 of the review document, a high percentage of responders is identified as having a high "within-group correlation?" The statement is unclear and needs elaboration to allow the reader to understand its significance.

On page 65, the discussion of individual versus group level would seem to be better explained as quantal versus continuous data. The rationale for the method used in categorical regression to model continuous data is not clear and seems inappropriate (The Committee did not understand the rationale). For example, in the categorical regression in Appendix C, 15 of 25 human volunteers were nonresponders to exposure to 7.0 mg/m^3 H_2S for 30 minutes. These data made the same contribution to the preferred analysis (m = 1) as if there had been only one subject. Likewise in experiment 21085, all animals died at dose of 695 mg/m^3, but the data were treated as if there were only a single animal.

One of the main advantages of categorical regression is that the concentration-duration plots can summarize data from different studies and endpoints. However, these plots are not clearly explained. Apparently a point is plotted for each severity observed in each group. Consequently, each group may have several corresponding points plotted, all at the same place on the graph. Are the points separated so that all can be seen? It would be more informative if the notation for censored observations were expanded to indicate the potential range of responses.

The H_2S example in Appendix B is not fully explained. For example, were concentration and duration log-transformed in the model? Why are data points different in Figures C-2 and C-3? In Figure C-3, wouldn't it be better to replace the censored indicators with the corresponding worst-case assumptions?

3.11 Additional Comments

On page 24, a statistical power calculation is based only on the experimental conditions, and does not reflect the experimental outcome. Simple confidence limits for the potential size of the response take into account the experimental results and are generally more useful than formal power calculations. For example, if a confidence interval includes zero, the study is negative. If, however, the upper confidence limit is (for example) 10 times the response found in a positive study, then the negative study had very low power.

On pages 50-51, the default assumptions for extrapolating to different durations seem reasonable. On the other hand, the relationship between the standard deviation and severity was unclear to the Committee.

The Committee recommends that the EPA specify a particular level of increased incidence for defining a BMC, e.g., 10%, and not have it depend upon the severity. This is apparently the approach adopted by EPA, as the document states on page 70 that "the 10% response had been adopted for dichotomous data", although this statement appears to be somewhat at odds with the discussion on page 53. Some of the EHC recommend uniform usage of a 10% response level to facilitate comparison of benchmark calculations for different substances. If desirable, severity could be still taken into account by explicitly applying a "severity factor", similar to other uncertainty or safety factors, in the analysis. Other members of the Committee are against uniform usage of a 10% response level.

On page 54, there is a discussion of the difficulties in applying benchmark modeling to continuous data, and two approaches are proposed. However, this discussion does not include the procedure proposed by Crump (1995). This procedure allows a single definition of effect to be used for different continuous endpoints. It also allows dichotomous and continuous data to be treated in a common manner. The EHC recommends this procedure to the EPA for consideration.

On page 56, line 25 "H is _a_ sigmoidal function..."

On page 57, line 1, here also, effect is confused with dose: s is an effect level; EL., as defined by EPA, is a dose.

On page 56, line 11, what is defined as the "effective exposure (concentration and duration)" is actually a duration only, corresponding to a given concentration and probability of an effect.

On page 58, lines 1 and 2, the statement indicates a lack of understanding of the statistical methodology. A likelihood can be generated for all types of data. Also, many "values" can be generated by categorical regression in addition to the likelihood.

On page 65, the Committee did not understand the application of categorical regression to continuous data. The methodology is questionable.

4. SUMMARY OF RECOMMENDATIONS

The EHC recommends that:

a)   The EPA should not include categorical regression in deriving the acute reference exposure methods until it can justify its use based on biological plausibility. Additional experience is needed applying the data sets to determine whether categorical regression adds anything above and beyond the NOAEL and benchmark approaches.

      (1)   The "g" factor, if still used, needs further explanation in the document. Also, the Agency should be able to present examples where the data fits the line.

      (2)   The Agency should also validate its assumption re categorical regression that all the probability curves for the various severities are parallel.

      (3)   The EPA should revisit its use of confidence intervals that are based on Wald-type statistical theory.

b)   The EPA should modify the expert system for categorizing severity to include clinicians, economists, and other individuals with applicable expertise.

c)   The Agency should explore its application of Haber's law in the ARE methodology in light of recent findings which indicate that Haber's law will be predictive when the time factor includes both calendar and pharmacokinetic time (i.e., the rate at which a chemical clears from the body).

d)   In applying Haber's law, the Agency should only extrapolate forward, not backwards, i.e., from shorter to longer exposure durations, but not from longer to shorter durations, since Haber's law does not hold if one tries to extrapolate (for example) from a 90-day study to a 6 hour exposure.

      1)   Limits should be placed on the extrapolation method so that it does not occur significantly outside of the dose response curve.

      2)   The EHC also found that the calculations are lacking in much the same way as the BMC and NOAEL/LOAEL calculations in defining risk to children. The Agency should explore modifying the methodology to address risk to children.

e)   The definitions of NOAEL and LOAEL on page 23 of the review document were incorrect and must be rewritten.

f)   The Inhalation Reference Concentrations (RfCs) and the Acute Reference
        Exposures (AREs) should be harmonized for the dosimetric adjustments.

g)   The Agency should reassess the database to determine the applicability of
        categorical regression and then return to the SAB for a follow-up review of the
        revised ARE methodology.

APPENDIX A - ACRONYMS AND ABBREVIATIONS

AEL	adverse effect level
ARE	acute reference exposure
BEC	biological equivalent concentration
BMC	benchmark concentration
BMD	benchmark dose/concentration
C	concentration
CAAA	Clean Air Act Amendments
CatReg	categorical regression
DAF	dosimetric adjustment factor
EHC	Environmental Health Committee
ET	extrathoracic (region from the nares to the trachea)
FOB	Functional Observation Battery
HEC	human equivalent concentration
IRIS	Integrated Risk Information System
K_m	apparent substrate affinity
LOAEL	lowest-observed-adverse-effect level
LOEL	lowest-observed-effect level
MACT	maximum achievable control technology
MLE	maximum likelihood estimate
NAS	National Academy of Science
NOAEL	no-observed-adverse-effect level
NOEL	no-observed-effect level
ORD	Office of Research and Development
PBPK	physiologically-based pharmacokinetic
ppm	parts per million
RDDR	regional deposition dose ratio
RGD	regional gas dose
RGDR	regional gas dose ratio
RT	respiratory tract
S.D.	standard deviation
T	time
TB	tracheobronchial region
UF	uncertainty factor

REFERENCES CITED

Allen, B.C. Kavlock, R.J. Kimmel, C.A. and E.M. Faustman. 1994a. Dose-response assessment for development toxicity: II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundamentals of Applied Toxicology*. Vol. 23, pp. 487-495.

Allen, B. C. Kavlock, R.J., Kimmel, C.A. and E.M. Faustman. 1994b. Dose-response assessment for developmental toxicity: III. Statistical models. *Fundamentals of Applied Toxicology*, Vol. 23, pp. 496-509.

Cox, D.R., and D. Oakes. 1984. Analysis of survival data. Monographs on statistics and Applied Probability. Chapman and Hall, London.

Crump, K. 1995. Calculation of benchmark doses from continuous data. *Risk Analysis*, Vol. 15, pp. 79-89.

Druckery, H. 1967. Potential Carcinogenic Hazards from Drugs in Evaluation of Risks (R. Truhaut, Ed. UCC Monograph Reviews, 7:60-77. Berlin, Springer Verlag.

EPA. 1998a. Methods for Exposure-Response Analysis for Acute Inhalation Exposure to Chemicals-Development of the Acute Reference Exposure, EPA/600/R-68/051, External Review Draft, Office of Research and Development, Washington, DC, April, 1998.

EPA. 1998b. *CatReg Software Documentation*, EPA/600/R-98/053, Review Draft, Office of Research and Development, Washington, DC, April 1998.

EPA. 1998c. CatReg Software User Manual, EPA/600/R-98/052, Review Draft, Office of Research and Development, Washington, DC, April 1998.

Guzelian, P.S., Henry, C.J., and S.S. Olin (eds.), 1992. Similarities and Differences between Children and Adults, Implications for Risk Assessment. ILSI Press, Washington, DC.

NAS. 1993. Pesticides in the Diets of Infants and Children. NRC, National Academy Press, Washington, DC.

SAP (Scientific Advisory Panel). 1998. A Set of Scientific Issues being Considered by the Agency in Connection with the use of the Food Quality Protection Act 10X Safety Factor to address the Special Sensitivity of Infants and Children to Pesticides. In: Final Report of the FIFRA Scientific Advisory Panel Open Meeting, Arlington VA, March 25, 1998, pp. 22-31.

U.S. Congress. 1990. *Clean Air Act Amendments of 1990*, PL 101-549, November 15, Washington, DC, U.S. Government Printing Office.

DISTRIBUTION LIST

Administrator
Deputy Administrator
Assistant Administrators
EPA Regional Administrators
Director, National Center for Environmental Assessment
Director, Office of Science Policy, ORD
EPA Laboratory Directors
EPA Headquarters Library
EPA Regional Libraries
EPA Laboratory Libraries
Library of Congress
National Technical Information Service
Congressional Research Service