September 30, 1998

EPA-SAB-EC-98-013

Honorable Carol M. Browner
Administrator
U.S. Environmental Protection Agency
4-1 M Street SW
Washington, DC 20460

RE: Review of the USEPA's Report to Congress on Residual Risk

Dear Ms. Browner:

The SAB's Residual Risk Subcommittee of the SAB Executive Committee conducted a peer review of the Agency's Residual Risk Report to Congress in Research Triangle Park on August 3, 1998. The review focused on five specific charge questions:

a) Has the Residual Risk Report to Congress (Report) properly interpreted and considered the technical advice from previous reports, including:

(1) The NRC's 1994 report "Science and Judgment in Risk Assessment"

(2) The 1997 report from the Commission on Risk Assessment and Risk Management in developing its risk assessment methodology and residual risk strategy?

b) Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources?

c) Does the Report provide an adequate characterization of the data needs for the risk assessment methods?

d) Does the Report provide adequate treatment of the inherent uncertainties associated with assessment of residual risks?

e) Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program?

The attached SAB consensus report provides an answer to each of these questions. In short, the SAB found the Report to be a generally good draft of a strategy document, but one that must be strengthened in a number of important places prior to submission to Congress. The Subcommittee was highly supportive of the Agency's coming back to the SAB in 1999 with examples in which the Report's strategy is used in specific cases.

Overall, the Report utilizes the risk assessment(RA)/risk management (RM) framework, endorsed by the SAB and others. It emphasizes the dynamic and evolving nature of the RA process by not being overly prescriptive, while also providing some bounds to the process in both the areas of RA and risk management RM. The Agency has clearly studied the National Research Council and Commission on RA/RM reports that related to this topic and has addressed many of the concerns and suggestions that they raised. At the same time, there are additional points that should be confronted more directly, including the following:

a) The Report gives a misleading impression that more can be delivered than is scientifically justifiable, given the data gaps and limited resources (e.g., time, funding) for conducting the residual risk assessments. The Subcommittee recommends that the Report more carefully convey the limitations of the data, models, and methods that are described or that would be needed to carry out the residual risk assessment activities. For example, a frank and clear discussion of: (a) current limitations in available methods (e.g., assessment of ecological risks at the regional ecosystem level) and data (e.g., emissions, IRIS, HEAST); (2) methods for reducing data gaps (e.g., the promise of uncertainty analysis to value-rank data gaps); and (3) priorities for research and management action should be provided.

b) The Report should contain or cite specific examples to clarify what some of the bold, but vague, language is intended to convey. Specific examples and/or citations of existing examples would clarify its discussion of the many complex and difficult issues involved.

c) There needs to be a more clearly described screening approach that will prioritize stressors for assessment and will conserve Agency resources. Unless the Agency carefully prioritizes its assessments and conserves its resources, the program could evolve either into a wide, but shallow, program that fails to adequately quantify and target residual risks or into a program that fails to address a sufficient number of pollutants and sources, due to over-analysis of just a few cases.

d) The Report should be more explicit about how the residual risk assessments will be used to make risk management decisions. If the intent is to increase the amount of

science that goes into risk reductions decisions, then it is useful in this strategy document to describe the interaction between the risk assessment and its application in the subsequent decisions that will need to be made as part of the risk management process.

Our report also provides a large number of other specific points of advice that the Agency should carefully consider.

We appreciate the opportunity to review the Report and look forward to the Agency's response and to future SAB review of the application of this strategy.

Sincerely,

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

/signed/
Dr. Philip Hopke, Chair
Residual Risk Subcommittee
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 Agency, 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 Residual Risk Subcommittee of the Science Advisory Board's (SAB) Executive Committee convened in public session on August 3, 1998 to review the U.S. Environmental Protection Agency's draft Residual Risk Report to Congress (Report). The Report describes the strategy methods the Agency will use to assess the risk remaining, (i.e., the residual risk) after maximum achievable control technology (MACT) standards, applicable to emissions sources of hazardous air pollutants (HAPs), have been promulgated under Section 112(d).

In short, the SAB found the Report to be a generally good draft of a strategy document, but one that must be strengthened in a number of important places prior to submission to Congress. The Subcommittee was highly supportive of the Agency's coming back to the SAB in 1999 with examples in which the Report's strategy is used in specific cases.

The SAB endorses the underlying the risk assessment (RA)/risk management (RM) approach described in the Report At the same time, there are additional points that should be confronted more directly and explicitly, including the following: a) The Report should more carefully convey the limitations of the data, models, and methods that are described or that would be needed to carry out the residual risk assessment activities; b) The Report should contain or cite specific examples to clarify what some of the bold, but vague, language is intended to convey; c) There needs to be a more clearly described screening approach that will prioritize stressors for assessment and will husband Agency resources; and d) The Report should be more explicit about how the residual risk assessments will be used to make risk management decisions.

The SAB report contains many other specific comments, as well as an appendix containing written comments from individual members.

Keywords: Residual Risk, hazardous air pollutants, HAPs, MACT, IRIS

SCIENCE ADVISORY BOARD
EXECUTIVE COMMITTEE
RESIDUAL RISK SUBCOMMITTEE

ROSTER

CHAIR
Dr. Philip Hopke, Department of Chemistry, Clarkson University, Potsdam, NY

SAB MEMBERS
Dr. Michele Medinsky, Chemical Industry Institute of Toxicology, Research Triangle Park, NC

SAB CONSULTANTS
Dr. Gregory Biddinger, Exxon Company-USA, Houston, TX

Dr. Thomas Burke, Johns Hopkins University, Baltimore, MD

Dr. H. Christopher Frey, North Carolina State University, Raleigh, NC

Mr. Thomas Gentile, New York State Dept of Environmental Conservation, Albany, NY

Dr. Warner North, NorthWorks Inc., Mountain View, CA

Dr. Gilbert Omenn, University of Michigan, Ann Arbor, MI

Dr. George E. Taylor, Jr., George Mason University, Fairfax, VA[1]

Dr. Rae Zimmerman, New York University, New York, NY

SAB STAFF
Dr. Donald Barnes

Ms. Priscilla Tillery-Gadson and Ms. Betty Fortune

[1] Affiliated with University of Nevada-Reno, Reno, NV at the time of the meeting.

TABLE OF CONTENTS

"2. Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources? *See especially Chapter 3, including discussions on health effects, dose- response, exposure, and ecological effects assessment. See also Chapter 4, screening and refined assessments (pp. 103-122).*

"3. Does the Report provide an adequate characterization of the data needs for the risk assessment methods? *See especially Chapter 3 (pp. 50-63) and Chapter 4 (pp. 103-122).*

"4. Does the Report provide adequate treatment of the inherent uncertainties associated with assessment of residual risks? *See especially Chapter 4 (pp. 89-95).*"

"5. Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program?"

This Appendix contains the final written comments from each of the Subcommittee members. These comments are included in this SAB Report so that the Agency and the public can a) benefit from the specific comments and b) appreciate the range of views represented on the Subcommittee. While all of these comments are commended to the Agency for their careful consideration, unless a comment is addressed explicitly in the body of this SAB Report, the comment should not be represented as the collective view of the Subcommittee.

1. EXECUTIVE SUMMARY

Section 112(f)(1) of the Clean Air Act (CAA), as amended, directs ERA to prepare a Residual Risk Report to Congress (Report) that describes the methods to be used to assess the risk remaining, (i.e., the residual risk) after maximum achievable control technology (MACT) standards, applicable to emissions sources of hazardous air pollutants (HAPs), have been promulgated under Section 112(d). The Report presents EPA's proposed strategy for dealing with the issue of residual risk and reflects consideration of technical recommendations in reports by the National Research Council ["Science and Judgment"] (NRC, 1994) and the Commission on Risk Assessment and Risk Management (CRARM, 1997). As a strategy document, the Agency's Report describes general directions, rather than prescribed procedures. The announced intent is to provide a clear indication of the Agency's plans while retaining sufficient flexibility that the program can incorporate changes in risk assessment methodologies that will evolve during the 10-year lifetime of the residual risk program.

In June, 1998, the Science Advisory Board (SAB) was asked to review the Agency's April 14, 1998 draft Report to Congress on Residual Risk. The Board was asked to focus primarily on the five specific charge questions that are addressed in the report:

1. Has the Residual Risk Report to Congress (Report) properly interpreted and considered the technical advice from previous reports,
   including:
   a. The NRC's 1994 report "Science and Judgment in Risk Assessment"
   b. The 1997 report from the Commission on Risk Assessment and Risk Management in developing its risk assessment methodology and residual risk strategy?
2. Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources?
3. Does the Report provide an adequate characterization of the data needs for the risk assessment methods?
4. Does the Report provide adequate treatment of the inherent uncertainties associated with assessment of residual risks?
5. Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program?

An SAB Subcommittee of the Executive Committee met in public session on August 3, 1998 at the USEPA main auditorium in Research Triangle Park, NC. Written comments prepared before and after the meeting by Subcommittee members form the basis for this report. Those comments are included in Appendix A for the edification of the Agency as an illustration of the issues identified by the Subcommittee members and the range of views expressed. A more detailed description of the SAB process for this review can be found in Appendix B.

In short, the SAB found the Report to be a generally good draft of a strategy document, but one that must be strengthened in a number of important places prior to its submission to Congress. The Subcommittee was highly supportive of the approach that the Agency described in terms of coming back to the SAB in 1999 with examples in which the Report's strategy is applied to specific cases.

Overall, the Report utilizes the risk assessment(RA)/risk management (RM) framework, endorsed by the SAB and others. It emphasizes the dynamic and evolving nature of the RA process by not being overly prescriptive, while also providing some bounds to the process in both the areas of RA and RM. The Agency has clearly studied the National Research Council and Commission on RA/RM reports that related to this topic and has addressed many of the concerns and suggestions that they raised. At the same time, there are additional points that should be confronted more directly, including the following:

1. *The Report gives a misleading impression that more can be delivered than is scientifically justifiable, given the data gaps and limited resources (e.g., time, funding) for conducting the residual risk assessments. The Subcommittee recommends that the Report more carefully convey the limitations of the data, models, and methods that are described or that would be needed to carry out the residual risk assessment activities.*

   The task of conducting so many assessments of the risks remaining after implementation of MACT controls is daunting, but doable. While the Report describes a general strategy for accomplishing this task, it does not address many of the outstanding, practical difficulties that will have to be overcome in carrying out the strategy. For example, there will likely be many situations in which the data implied in the strategy are absent. Although a number of options exist, it is not clear what the Agency will do in such cases. Other problems that need attention in the near term include: computer models that have had only limited independent testing for their application to a particular problem and/or have not been adequately validated for its general applicability across a wider array of situations, information in important toxicological databases that is outdated or has had limited peer review, and special limitations in information and tools for ecological risk assessment. The Congress and the public, on the basis of reading this Report, may have unrealistically high expectations of what the Agency can, in fact, deliver in terms of the accuracy, precision, and timeliness of residual risk assessments.

2. *The Report should contain or cite specific examples to clarify what some of the bold, but vague, language is intended to convey.*

   The Report lacks any specific examples and/or citations to existing examples to illustrate its discussion of the many complex and difficult issues involved, such as, but not limited to, the following:

   a. Involving stakeholders in the process, which is particularly important when it comes to sharing information among the Federal and State Governments and industry..

b. Determining the criteria for when to use other than default assumptions.
c. Addressing background contamination and competing sources of risks (e.g. mobile and area sources). .
d. Dealing with the trade-off between risks from HAPS and possible risks posed by measures to reduce the HAPS risks.
e. Assessing risks in the face of significant limitations in the available data, the lack of validation of existing and emerging computer models, and the need to consider uncertainty in the results.
f. Employing screening tiers and emerging risk assessment methodologies in such a way that scarce resources are targeted on the most important assessments and are not expended on resource-intensive, low-information-yield analyses.
g. Providing a public health perspective to these issues.

3. *There needs to be a more clearly described screening approach that will prioritize stressors for assessment and will conserve Agency resources. The Report should more clearly present the approach by which the Agency will perform the screening and prioritization.*

There is the potential that the Residual Risk program could evolve into a large, resource-intensive activity unless there is an appropriate and well-supported screening approach in place to prioritize assessments among the 188 pollutants and 174 source categories. The screening methods should be such that they avoid generating a large number of "false positives" -- that would drain scarce RA resources -- or "false negatives" -- that could result in leaving high risk situations unaddressed. Unless the Agency carefully prioritizes its assessments and conserves its resources, the program could evolve either into a wide, but shallow, program that fails to adequately quantify and target residual risks or into a program that fails to address a sufficient number of pollutants and sources, due to over-analysis of just a few cases.

4. *The Report should be more explicit about how the residual risk assessments will be used to make risk management decisions.*

The Subcommittee recognizes that the Report is a description of a strategy for RA, not for RM, per se. However, as S&J and the CRARM report each emphasize, there should be open communication between risk assessors and risk managers at the beginning of the process, so that it is clear how the RA will fit into the RM process. If the Residual Risk program is, indeed, to be "science-based", then it is important that there be, even in a strategy document, some discussion of what type of RA is needed and how its results will be factored with other legitimate risk management factors during the final stages of decision making.

The Subcommittee strongly encourages the Agency to implement their plan to bring to the SAB for review in 1999 some applications of the Residual Risk strategy as specific illustrations of how these complex issues will be addressed. This approach will permit more detailed discussion

of many of the implementation issues that members felt will arise when residual risk assessments are made.

Considering a larger issue beyond its specific Charge, the Subcommittee expressed some concern about the manner in which risks from HAPs are being addressed, when compared with the risks posed by Section 109 Criteria Air Pollutants (CAPs). There are differences in the wording of the Clean Air Act Amendments as to the level of risk avoidance that should be provided. This incongruity is puzzling and suggests that it may be useful to reevaluate how risks are assessed and managed for these two types of airborne pollutants. We recognize that the current legislation requires that these two classes of pollutants be treated separately. However, since the Agency was specifically asked to suggest changes in the legislation, there is an opportunity to propose a more comprehensive framework upon which to build the assessment and management of the risks from both HAPs and CAPs. Such a broader public health perspective would result in greater improvements in health and environmental benefits for a given expenditure of resources. The Agency has taken some steps towards a comprehensive view of HAPs and CAPs in its Report to Congress on the Costs and Benefits of the Clean Air Act, 1970-1990 (USEPA, 1997) that has been reviewed earlier by the SAB (SAB, 1997; SAB, 1996) and those steps should be continued. The contrast in relative benefits of the two programs was revealing.

In addition, the Agency Staff should consider outlining a number of the most important Residual Risk issues in a policy memo to top management; e.g., the limitations on what science can deliver and the comparison between the Section 112 (HAPs) program and the Section 109 (CAPs) program. These managers should be made aware of the problems involved and be given the opportunity to provide the kind of guidance that would clarify these matters for the benefit of those both inside and outside of the Agency.

In summary, the Agency's Report is a useful strategic document that will help guide the Agency as it moves ahead with the Residual Risk program. However, the Subcommittee recommends that the Agency be more candid with Congress and the public about what can be accomplished with existing limitations in data, models, methods, time, and resources. The Subcommittee has pointed out many areas that will require more thought, more documentation, and more articulation before the program is actually implemented.

2.0 INTRODUCTION

2.1 Background

Section 112(f)(1) of the Clean Air Act (CAA), as amended, directs EPA to prepare a Residual Risk Report to Congress (Report) that describes the methods to be used to assess the risk remaining, (i.e., the residual risk) after maximum achievable control technology (MACT) standards, applicable to emissions sources of hazardous air pollutants (HAPs), have been promulgated under Section 112(d). The Report presents EPA's proposed strategy for dealing with the issue of residual risk and reflects consideration of technical recommendations in reports by the National Research Council ["Science and Judgment"] (NRC, 1994) and the Commission on Risk Assessment and Risk Management (CRARM, 1997). As a strategy document, the Agency's Report describes general directions, rather than prescribed procedures. The announced intent is to provide a clear indication of the Agency's plans while retaining sufficient flexibility that the program can incorporate changes in risk assessment methodologies that will evolve in the future.

2.2 Charge

In June, 1998, the Science Advisory Board (SAB) was asked to review the Agency's April 14, 1998 draft Report to Congress on Residual Risk. The Board was asked to focus primarily on the following five specific charge questions:

1. Has the Residual Risk Report to Congress (Report) properly interpreted and considered the *technical* advice from previous reports, including:
    a. The NRC's 1994 report "Science and Judgment in Risk Assessment" (*see especially pp. 8-11 of the Residual Risk Report and the Executive Summary, pp. 1-15, from the NRC report*); and
    b. The 1997 report from the Commission on Risk Assessment and Risk Management (*see especially pp. 11-15 from the Residual Risk Report and the CRARM Report's discussion on "Tiered Scheme for Determining and Managing Residual Risks" on pages 109-112*), in developing its risk assessment methodology and residual risk strategy?

2. Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources? *See especially Chapter 3, including discussions on health effects, dose-response, exposure, and ecological effects assessment. See also Chapter 4, screening and refined assessments (pp. 103-122).*

3. Does the Report provide an adequate characterization of the data needs for the risk assessment methods? *See especially Chapter 3 (pp. 50-63) and Chapter 4 (pp. 103-122).*

4. Does the Report provide adequate treatment of the inherent uncertainties associated with assessment of residual risks? *See especially Chapter 4 (pp. 89-95).*

5. Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program?

2.3 SAB Review Process

The SAB Subcommittee was recruited following nominations received from SAB Members and Consultants, the Agency, and outside organizations. The Subcommittee met in public session on August 3, 1998 at the USEPA main auditorium in Research Triangle Park, NC. Written comments prepared before and after the meeting by Subcommittee members form the basis for this report. Those comments are included in Appendix A for the edification of the Agency and the public as an illustration of the issues identified by the Subcommittee members and the range of views expressed. A more detailed description of the SAB process for this review can be found in Appendix B.

3.0 RESPONSES TO SPECIFIC CHARGE QUESTIONS

3.1 Charge Element 1

1. Has the Residual Risk Report to Congress (Report) properly interpreted and considered the technical advice from previous reports, including:
   a. The NRC's 1994 report "Science and Judgment in Risk Assessment" (S&J) (*see especially pp. 8-11 of the Residual Risk Report and the Executive Summary, pp. 1-15, from S&J*)

Overall, the draft Report is responsive to the recommendations in the 1994 National Research Council Report, Science and Judgment in Risk Assessment (S&J) (NRC, 1994). The comments below and in Appendix A are intended to help in the process of refining and improving the current draft Report, which is "a work in progress" leading to developments of an extraordinarily complex regulatory program that will shape air pollution policy for decades to come. The Report describes a strategy that integrates a broad range of public health, regulatory, technical, and social considerations to provide a framework for implementing the Act. Many of the Subcommittee's comments the follow are motivated by a desire to see main themes touched on in this draft Report or in S&J set forth at greater length or with greater clarity. Other Subcommittee comments address details of implementation and the need to go even further toward a flexible, iterative, and tiered system of the type described in S&J.

Perhaps the most important need is to explain to Congress the large uncertainties and judgmental basis for cancer risk numbers in default assumptions, such as low-dose linearity, and the importance of these issues for risk assessment. (See S&J, Executive Summary, first and third bullet at top of page 10, and Appendix B of the draft report, page B-3, first new paragraph; as well as the extensive discussions in the two-volume CRARM report.)

It is particularly important to acknowledge the uncertainty regarding whether the dose-response relationship for carcinogens (and some non-carcinogens) at low doses is linear or nonlinear. This uncertainty in low-dose linearity is going to be critical for many of the regulatory decisions on HAPs. The uncertainty and the underlying science should be clearly explained to decision makers and Congress, and not masked in discussion of complex risk assessment procedures, such as benchmark dose and the linearized multistage model. The discussion should be transparent and readily accessible to the non-risk specialist.

More attention should be paid to the S&J recommendation that the Agency improve its criteria for defaults and for departure for defaults. While the defaults issue is mentioned on page 10 of the draft Report, it is not developed adequately. A reader from Congress unfamiliar with cancer risk assessment might not even know what the National Research Council was talking about with regard to defaults, since the concept of a default option is not introduced and explained. This issue is discussed at length in S&J and motivates some of its most important recommendations in Chapters 6 and 12 of that report. According to S&J, such defaults should be

noted and explained. Exceptions should be made in those cases where an adequate scientific basis exists. For example, the Agency has taken positions on excluding certain rat kidney tumors and thyroid tumors from consideration (USEPA, 1991; USEPA, 1997). The Agency concluded from scientific data that the results from animal studies do not indicate the potential for human disease because different biological mechanisms are involved in the different species.

Case studies would be very useful devices for demonstrating how an iterative, tiered process actually works. In fact, the Report's use of the benzene decision (Report Appendix B) is helpful in this regard. However, each of the comments received from the public at the meeting and all of the Subcommittee members agreed that the Report would be helped by referring to additional cases studies in order to clarify the Report's often too general language on how the Agency intends to address some of the most difficult issues in risk assessment identified in S&J; e.g., stakeholder involvement; an iterative, tiered scheme for assessment; and introduction of other than default assumptions. S&J provides some guidance on these matters. For example, S&J contains several useful case studies in Chapter 6 and in its appendices F and G. These and other case studies (see Paustenbach, 1989, and publications in *Risk Analysis: An International Journal*) should be cited. In addition, examples could be drawn from the experience of the Agency or State or Local Air Toxics Agencies that have conducted risk assessments on a specific source category; e.g. Municipal Waste Combustion Facilities, and made the subsequent risk management decisions about the significance of the remaining risk.

The Agency plans to rely extensively on the Integrated Risk Data System (IRIS) [as well as the Health Effects Assessment Summary Tables (HEAST) system] in conducting the residual risk analyses. The substantial limitations of the IRIS data, in terms of outdated information or information that has had limited peer review, have been explicitly discussed in S&J; cf., chapter 12, pp 250-1, 265. The Report should address those limitations and acknowledge the importance of providing higher quality in this data base through adequate financial support and appropriate internal and external peer review. The Agency needs to ensure adequate quality of all of the data in IRIS, as well as an expansion of the data base to become a risk assessment data base (including ecological risk), not just a toxicology data base. As it stands, the Agency continues to be criticized for failure to provide adequate resources for IRIS (Risk Policy Report, 1998).

In a related matter, the Agency should explore the mechanisms for sharing and using quality data that may exist beyond the confines of IRIS and HEAST. Sources of such data may include other Federal agencies (e.g., the Agency for Toxic Substances and Disease Registry), State Governments (e.g., the State of California's assessments of HAPs by its Office of Environmental Health Hazard Assessment) and industry.

There should be greater emphasis on setting priorities for research and further data collection, as an output from the iterative, tiered approach. The statutory need for residual risk assessments under Section 112 should provide motivation not only for the Agency, but also for industry and other government agencies (e.g., National Institute of Environmental Health Sciences (NIEHS)) to conduct additional needed research and data collection. Again, S&J is

quoted on page 10 and Exhibit 1, reproducing the S&J figure that derives from Figure 1 in the first National Research Council report on risk assessment (NRC, 1983), but the ideas are not developed. Ideally, the Report would describe the current public and private research agendas, timetables, and how the Agency will be assembling and evaluating information collected under other statues, such as the Toxics Substances Control Act (TSCA) to fill the data gaps associated with potential health and environmental effects of individual HAPs and HAPs mixtures. (See Appendix A.4 for the potential utility to four State environmental agencies sharing information collected under TSCA.) Sharing of information between the Federal and State Governments and industry will be important to the success of the Residual Risk program.

The Agency should consider convening a workshop to review of the recommendations of the S&J report and their applicability to ecological risk assessment (eco RA) . There are numerous S&J recommendations that are applicable to ecological RA. A conscious effort -- involving both health and ecological scientists -- to do so would help to integrate human health and ecological risk assessments conceptually, at first, and practically, later. It is interesting to note that there was an earlier workshop, under the auspices of the NRC, to examine ecological RA in connection with the NRC's 1983 report on risk assessment (NRC, 1993).

b. The 1997 report from the Commission on Risk Assessment and Risk Management (*see especially pp. 11-15 from the Residual Risk Report and the CRARM Report's discussion on "Tiered Scheme for Determining and Managing Residual Risks" on pages 109-112*), in developing its risk assessment methodology and residual risk strategy?

The Report is heavily influenced by the recommendations of CRARM and, for the most part, does an effective job of integrating its recommendations into the framework. Specifically, the description of the CRARM reports appropriately emphasizes the risk management framework, the engagement of stakeholders, the early effort to put problems into a public health and ecologic context, and the need to move from one chemical, one medium, one risk at a time to multi-source, multi-media, multi-chemical, and multi-risk analysis and management. Such contexts should be an explicit part of this residual risk strategy.

In order to demonstrate that each of the CRARM recommendations were considered, it is useful that Section 5.3.5. of the Report lists them all and describes how they were addressed, even though the descriptors in the table are necessarily terse. For the most part the Agency is "in the process" of developing strategies to address each point. Potential differences in implementation and interpretation are listed in the following subsections:

3.1.1 "Characterize and articulate the scope of the national, regional, and local air toxics problems and their public health and environmental contexts." (USEPA, 1998, p. 111).

The entire Agency is obviously just beginning this process, particularly in regard to the

public health and environmental contexts. The CRARM calls for a broad public health approach which examines the actual health impacts on the affected communities, and considers the residual risk in the context of the health status of the population. While the Report mentions the collection of population health data and potential integration of epidemiological approaches, no specific methods are detailed, and no commitment is made to tracking the health status of the population. The proposed program is largely driven by animal-based cancer bioassays to estimate the public health context. Without developing a more detailed approach, it is not possible to determine if the Agency is actually implementing this CRARM recommendation. More importantly, without a specific strategy to evaluate population health status, it will be difficult -- if not impossible -- to determine if the residual risk management makes any difference in the public's health. Thus, a well-articulated approach might also be useful in demonstrating achievement in preventing adverse health outcomes.

3.1.2 "At facilities that have upper bound cancer risks greater than one in 100,000 persons exposed or that have concentrations greater than reference standards, examine and choose risk reduction options in light of total facility risks and public health context." (USEPA, 1998, p. 111)

According to the CRARM Report, this recommendation is intended to result in development of a flexible bright line that considers local public health impacts and the total facility risk. The Agency does not adopt the one in 100,000 approach, opting instead for the flexible approach of the benzene NESHAP. This issue should be addressed in greater detail in the Report in order to better represent the recommendation of CRARM, which was a publicly-aired proposal for using 10^{-5} as the "bright line" for action after refined risk assessments, rather than the extremely conservative 10^{-6} (both upper bound risk estimates). In fact, the 10^{-6} is not proposed for each chemical, but 10^{-6} for the combined effects of all carcinogenic HAPs that may be emitted by a source. Thus, some information from the MACT experience should be inserted to indicate the number of HAPs per source category and their carcinogenic potential. In short, the Report should more clearly articulate the Agency's consideration and disposition of this CRARM recommendation.

It is not apparent just how the Agency will interpret "public health context". An example or a flowchart would be helpful in clarifying the Agency's discussion in Section 4.1.1 of the Report, on p. 65 and following.

3.1.3 "Consider reduction of residual risks from source categories of lesser priority." (USEPA, 1998, p. 112)

The Agency interprets this as a mandate to do the "worst first", and considers the Report to address this recommendation. Further consideration of lesser sources should be included in order to address the management of high background risks, to protect populations with high aggregate or cumulative risk, or to consider the public health of sensitive populations.

3.1.4 Contrasting response to CAPs and HAPs

Although the Subcommittee was not asked to comment on the Agency's conclusion that no legislative changes are recommended, the Subcommittee feels compelled to provide a technical perspective on this policy decision. Specifically, the Subcommittee believes that the Agency should work with the Congress and the various constituencies to reconsider the peculiar and now-dated distinction between CAPs and HAPs. As one small example, there is an "adequate" margin for CAPs for which NAAQS are generated to protect the entire U.S. population and the "ample" margin for Section 112 HAPs to which much more limited portions of the population are actually exposed. Also, in 1970 there was an overwhelming preoccupation with cancer risks, and a general desire to reduce risks to zero; there was little attention to other life-threatening, serious, salient adverse health effects. We know better now, yet, as the CRARM points out, we still have a long way to go in applying comparable analysis and risk management approaches to section 109 and section 112 pollutants. The Agency has taken some steps towards a comprehensive view of HAPs and CAPs in its Report to Congress on the Costs and Benefits of the Clean Air Act, 1970-1990 (USEPA, 1997) that has been reviewed earlier by the SAB (SAB, 1997; SAB, 1996) and those steps should be continued. The contrast in relative benefits of the two programs was revealing.

3.1.5 Continued use of extreme exposure and risk scenarios

The CRARM report, as well as the Agency's Risk Characterization Guidance, emphasizes the use of "high end" (e.g., 90% percentile exposures), rather than the more extreme Maximum Individual Risk (MIR) and Maximum Exposed Individual (MEI) concepts are featured in the Report. Although the MIR is proposed to be used only in upper-bound screening studies, the Subcommittee would like to emphasize that more complete analyses should be made using the approaches outlined by CRARM (CRARM, Vol 2, p. 74). This process would involve using a more realistic individual exposure estimate (e.g., EPA’s high end exposure estimate or a maximally exposed actual person), coupled with the estimates of total number of potentially exposed individuals. Subsequent risk management decisions would be based on refined iterations of the exposure assessment that evaluate the distribution of a population’s varied exposures, examining any segments of the population that have unusually high exposures or unusually high susceptibility.

The Report also proposes to continue the Agency's practice of using the 10^{-6} upper bound as the individual risk level that generally meets the "ample margin of safety" criterion, rather than the 10^{-5} level chosen and recommended by the CRARM after extensive discussion in public hearings. The "margin-of-exposure"(MOE) analyses will likely show how remarkably conservative even 10^{-5} upper bound levels are, compared with other important health risks regulated by the Agency. The Report should clearly state the rationale for not following the CRARM recommendation on risk level.

3.1.6 Other CRARM-related topics

The CRARM, the S&J, and the Agency have all emphasized the importance of mode-of-action information in identifying hazards. The Report is curiously mute on this topic. It is understood that this information is not available for all HAPs, but again this data gap should be acknowledged and addressed appropriately.

The Risk Commission worked hard on the matter of mixtures and additivity. In general, they concluded that additivity was a highly conservative assumption; in many cases, related chemicals will be competing against each other for access to a common receptor or other target molecule. Again a better description of how the problem of mixtures will be handled relative to the CRARM discussion should be presented.

3.1.7 Other RA/RM extensions

The Agency should recognize that there are more paradigms for risk assessment/risk management that just those suggested in S&J and CRARM. For example, the Agency's Office of Research and Development (ORD) has developed and employed a strategic plan paradigm that, in some ways, adapts these other two to the principal research entity in the Agency (USEPA, 1996). This is important, because the Report needs to come to terms with how it provides for the integration of stakeholders; i.e., in providing data, in decision-making, etc. Finally, the National Research Council, "Understanding Risk" (NRC, 1996, p. 28) report implies still a fourth model that is far more interactive and involves stakeholders to a greater degree than any of the others. These other approaches should be acknowledged and a more complete description of the integration among them should be presented so that it is clear exactly how the process will work.

3.2 Charge Elements 2-4

2. Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources? *See especially Chapter 3, including discussions on health effects, dose-response, exposure, and ecological effects assessment. See also Chapter 4, screening and refined assessments (pp. 103-122).*
3. Does the Report provide an adequate characterization of the data needs for the risk assessment methods? *See especially Chapter 3 (pp. 50-63) and Chapter 4 (pp. 103-122).*
4. Does the Report provide adequate treatment of the inherent uncertainties associated with assessment of residual risks? *See especially Chapter 4 (pp. 89-95).*

Again, the reader is referred to written comments from the individual Subcommittee members (See Appendix A) to gain the full richness of the issues and opinions addressed by the Subcommittee members. While it is this report, *per se*, that represents the consensus position of the SAB, the individual opinions contain additional insights and perspectives that can be usefully considered by the Agency.

3.2.1 Health

3.2.1.1 General Comments

The Agency has developed a well-written, clear Report that outlines a very ambitious strategy for assessing human health residual risks as mandated by the Clean Air Act. However, assessment of such residual risks for a broad spectrum of endpoints as a result of exposure to mixtures of chemicals arising from multiple pathways is a daunting task, and a clearer description of the difficulties involved would provide a useful perspective on what can and cannot be accomplished.

3.2.1.2 Charge Element 2

The Agency's already daunting task is made more difficult by the following three model/data-related issues:

a. Many of the methods proposed by the Agency to assess these risks are in the development stage, even in the application to single chemicals.

b. Our toxicology knowledge of complex issues, such as the potential additive or interactive effects of chemical mixtures at low doses and the modes or mechanisms of action of the individual HAPs, is incomplete or rudimentary.

c. The data base for developing and validating models and assessing toxic effects is incomplete or absent for many HAPs.

Communicating the limits of our knowledge and risk assessment tools to Congress in this Report is essential in order to prevent the misconception that we know more than we do. Congress and the public should not place an inappropriate level of confidence on the accuracy and precision of the results of the residual risk analyses in light of the current limitations in the methods and available data.

Because of the complexity and comprehensiveness of this risk assessments, the Agency has appropriately elected to conduct the assessments in stages using a tiered iterative approach. Screening assessments will be used first. This approach will conserve limited human resources. However, it is important that all stakeholders be aware of the conservative, screening nature of the lower tier assessments that are designed to yield a certain level of false positives. Otherwise, there could be significant misinterpretation of the results.

The Agency presents a picture of the residual risk assessment process in broad brush strokes, as almost an idealized view of the process in which the implicit assumptions that undergird modeling strategies are correct, any and all data gaps are filled, and knowledge of mechanisms and modes of action is complete. However, the actual situation is much more

complex, and many unknowns are subsumed into the details. In short, translation of the principles, as laid out in this Report, into practice for the various individual risk assessments will be fraught with unknowns. It is incumbent upon the Agency to present these unknowns in a thorough, straightforward manner and acknowledge that these problems exist.

The Report should clarify that the need for risk assessments for acute non-cancer risks is related to the averaging times dictated in various regulations; e.g., "annual average concentration". The problem is, of course, that such measures of pseudo-chronic concentrations could be met by a few episodes of high intensity emissions connected by extended periods of low or zero emissions.

The Agency's approach to addressing acute exposures is still in draft (USEPA, 1998); an example of an important risk assessment methodology that is not yet in place. As the acute exposure document is being completed, the Agency should harmonize that approach, to the extent applicable, with the Agency approach to assessing non-cancer effects due to chronic exposure; i.e., Reference Concentration (RfC) methodology (USEPA, 1994; SAB, 1998a; SAB, 1991). For example, the dosimetric adjustments described in the documents are different at this point in time. Since both methodologies are assessing noncancer health effects, even though the toxic endpoints might be different, it is logical that both documents should describe similar dosimetric adjustments.

A second issue regarding risk assessments of acute health effects relates to the usefulness of categorical regression for setting points of departure for acute effects. The discussion in the Report (page 29) is an excellent example of the theoretical, idealized character of the Report. That is, the description of the plan of action and overview of the general concepts underlying categorical regression would draw little criticism. However, the reality is that the specifics of the methodology have not been widely accepted, nor is it likely that the data bases for many HAPs will be sufficiently robust to implement this methodology in more than a few instances (SAB, 1998b). These limitations to the implementation of this methodology are not provided.

In the discussion of chronic non-cancer effects, the Agency notes on page 27 the use of the Benchmark Dose (BMD) approach as an alternative to NOAEL to determine a dose without appreciable effect, based on experimental data. The Agency's acceptance of the BMD methodology is a very positive step forward. However, there is unresolved ambiguity about how the Agency will use the approach. For example, in other documents, the Agency has discussed a variety of levels that could be used as the BMD; e.g., point estimate or lower bound estimate on the 5% effective dose (ED05) or ED10. Further, most recently the Agency has introduced an additional uncertainty factor (UF), beyond the uncertainties employed in the traditional NOAEL/UF formulation, based on the judgment that the BMD is a finite response level and therefore more equivalent to a LOAEL than to a NOAEL; hence, the need for an additional UF. However, this additional UF is not universally accepted as appropriate in such cases. Because of the Agency's plans for the BMD are still in a state of flux, it is impossible to comment on the scientific basis of the application of the BMD in the case of residual risk analysis.

In its discussion of cancer effects on page 30 the Agency notes that "If animal data are used in the dose-response assessment, a scaling factor based on the surface area of the test animals relative to humans is used to calculate a human equivalent dose. Surface area is used for this scaling because it is a good indicator of relative metabolic rate." However, differences in the rates at which humans and laboratory animals metabolize xenobiotic chemicals (including many of the HAPs) do not always correlate with basal metabolic rate, and by extension the surface area scaling factor (Csanady et al, 1992; Seaton et al , 1994; Seaton et al, 1995). Thus, surface area may not always be a good indicator of the effective dose for chemicals that are metabolically activated. This scaling factor should really be referred to as the default value that is used in the absence of chemical-specific data.

The assessment of risks from chemical mixtures is another instance in which the Report cites an Agency methodology that is undergoing perhaps significant change. Both the generality of the specification of sources and the 17 categories of HAPs pose challenges that go beyond the capabilities of most traditional, single stressor-oriented risk assessment approaches. Commendably, the Agency is revising its Chemical Mixtures Risk Assessment Guidelines, first published in 1986. Therefore, it is not known at this time whether or how the procedures for assessing risks of mixtures will change significantly. The Agency's proposal on page 61 to calculate a Hazard Index "for all components of a mixture that affect the same target organ using the RfC (even if the RfC was derived based on an effect in a different target organ)" is confusing and requires further explanation. As stated, it appears that an RfC based on a lung effect, for example, could be combined with an RfC based on another organ effect, such as liver toxicity, to obtain the Hazard Index, a scientifically dubious proposal. An example of how this index would be applied in a specific case would be useful.

This concern about how the Agency plans to assess risks from mixtures was heightened by the bald statement (page 62) that "general additivity would include addition of effects that occur in different target tissues or by different mechanisms of action." The Report should make it very clear that such an approach of combining the effects of different chemicals acting by different mechanisms of action on different target organs is a policy decision, explicitly designed to generate an excess estimate of risk for screening purposes, is not based on science, and is not consistent with the Agency's existing guidelines for the best method of assessing risks from mixtures. According to science and the Agency's guidelines, additivity should be based on consideration of commonalty of mechanism; if chemicals do not act through a common mechanism their risks should be considered independently. This dependence on common mode of action for aggregating risks should apply to both cancer and non-cancer endpoints.

The Agency has made significant progress in examining the relevance of animal data to methods for assessing human health risk. These contributions, which should be mentioned in the Report, include

a. Rodent carcinogens that are not relevant for human assessment
b. Chemicals that are suitable for nonlinear analysis
c. Chemicals that are suitable for assessment via a Margin-of-exposure (MOE) approach

It would be instructive, for example, to identity which of the HAPs fall into any of these three groups.

Commenting in a context broader than the residual risk program, the Subcommittee recommends that the entire Agency seriously compare its whole philosophy and methods for protecting public health with the approach that has evolved in the public health community (PHC) over the past century. An example would be addressing childhood asthma. The public health context would be to reduce the incidence irrespective of source (power plants emissions vs indoor air pollution), of media (air vs water), of stressor (particulate matter (PM) vs microbes), of route of exposure (inhalation of air vs ingestion of food). In contrast, EPA in part because of its legislative mandates, tends to focus on one stressor (e.g., PM), in one medium (e.g., air), in one class of sources (e.g., stationary combustion), and one route of exposure (e.g., inhalation). PHC methods would provide an interesting and perhaps instructive alternative to the Agency's approach that has evolved over the past three decades and might provide a basis for more integrated strategies to reduce public health risks. Such an approach will be elaborated upon in a pending report from the SAB's Integrated Risk Project (SAB, 1998c).

Also, the Agency should build its residual risk program as a natural extension of the MACT program, benefiting from the experience gained from the efforts already underway. Therefore, a close monitoring and analysis of the results of the MACT program will provide insights on improving methods for estimating, projecting, and demonstrating emissions reductions, exposure reductions, and, over time, risk (endpoint) reductions in the range of greatest benefit and most reasonable cost.

Similarly, the Report encourage that the Agency to investigate more closely those state and local air toxics programs, with their associated analytic and methodological approaches, that are grappling with residual risk-related problems of their own. By keeping informed about state and local approaches, the Agency stands to improve over time its assessment and management methods for dealing with these 188 pollutants x 174 sources, with the end result of being respectful of the limited resources available for studies, analyses, and decision-making.

3.2.1.3 Charge Elements 3 & 4

The quality, accuracy, and completeness of the risk assessments will depend upon the quality, accuracy, and completeness of the data used in the risk assessments. The Agency should expand significantly on the issue of the data needs for conduct of the residual risk assessments and acknowledge the widespread data limitations. Limited data combined with default assumptions can result in risk assessments that are not well informed and that extend well beyond the boundaries of the underlying science.

In short, the Report should convey the limitations, data collection needs, and research needs associated with risk assessment both in evolving improved risk assessment methods and in providing the critical data needed to apply any methodology. It should give clearer context for

the current state of practice of risk assessment and provide a road map for what additional information, data, models, etc. would be needed to fully comply with the current requirements of the Clean Air Act regarding residual risk assessments. Where requirements of the Act appear to be optimistic or unrealistic, the Agency should give an indication of data collection and research activities that are needed in order to proceed with the conduct of the assessments. Agency policymakers need to better appreciate this disconnect between what is desired and what is possible in light of the limitations described above.

The current draft is too limited in its discussion of the problem. For example, in the Executive Summary of the Report, the Agency notes that "Information available on actual health effects resulting from exposures to air toxics is limited." As implied above, this thought should be expanded upon here in order to give Congress a fuller understanding of the implications of this important statement of fact for the results that will be derived in the Residual Risk program.

Also, to address this problem, the Report refers to a concerted effort to evaluate other types of human health data for possible correlations between exposure and adverse health effects, accessing such resources as disease registries, hospital and other medical records, morbidity reports, and incident/complaint reports at the State level. While this type of information is valuable for making sound public health decisions, the Agency should inform Congress that the nationwide compilation of such data is a major task, one that again illustrates the importance of close cooperation between the Agency, other Federal agencies, the States, and other stakeholders.

One approach to dealing with data gaps is through a progression of linked data sets (Zimmerman, 1990). For example, the most specific data for an air toxic risk assessment would be knowledge of a known health effect associated with a known exposure. If that information is unavailable, one works back to exposure indicators. If exposure information is unavailable, one then draws on source-based measures, etc.

Elsewhere throughout the Report where uncertainty is mentioned (e.g., the "Sources of Information for Hazard Identification" Figure on page 22), the Agency should be much more direct and thorough in explaining to the reader the extent of the data gaps and the consequences that they portend, in terms of both the magnitude of the uncertainties associated with the risks and the level of confidence in its overall risk assessments.

In fact, the impact of the data quality on the confidence associated with a guidance level has been addressed previously in the Agency's RfC guidelines, where a descriptor is given for the confidence in the data base. The Agency could use that discussion as a starting point for text in this Report that would inform the reader as to the limitations of the Residual Risk strategy in practice.

Because of the fundamental nature of the problem of data gaps, this issue should be highlighted in a separately identified section. A good starting point for the development of a section on "Data Base Limitations" might be a table listing the current HAPs and some assessment

as to the completeness of the toxicity data base for each of these chemicals; cf., table 6 in Appendix A of S&J on page 334 and an Agency effort developed in the same timeframe (USEPA, 1993). The type of information what could be displayed in such a table are the following:

a. Are there adequate chronic studies for assessing carcinogenicity, developmental and reproductive toxicity, and neurotoxicity?
b. Are there any structure-activity indications that a chemical might have toxic effects that would not be manifest in conventional toxicity studies, due to the lack of sensitivity towards these endpoints; e.g., immunotoxicity and/or respiratory tract hyperreactivity?
c. Even for chemicals for which there are sufficient data to justify a chemical's classification as a carcinogen; are there sufficient data to determine the mechanism or mode of action?
d. Are there sufficient data to provide mode of action information for all the HAPs for all toxicity endpoints that could be used in aggregating risks for determining residual risks from mixtures?

Such a table would enlighten Congress and others as to the difficulty of the Residual Risk task and the potentially large uncertainties associated with producing quantitative estimates. Further, that table might well stimulate stakeholders to generate pertinent, reliable data from new studies or bring forward such data from existing studies, such as those done for other regulatory requirements; e.g., TSCA. In this regard, the Report should describe some mechanism by which such valuable new data could be brought to the attention of the Agency for inclusion. The introduction of new data from interested and affected parties via an iterative, tiered, stakeholder-involvement process is one of the features recommended by the NRC reports (NRC, 1994; NRC 1996), and the Agency should be clear how it plans to develop and conduct such a process.

Section 4.2.3 of the Report neglects some broader sources of uncertainty; e.g.,

a. Uncertainty in selection of representative scenarios, including pollutant sources, transport, exposure pathways, exposed populations, etc.,
b. Uncertainty in the structure of models used to represent a given scenario, and
c. Uncertainty and variability in the inputs to the model(s).

The Report tends to focus only on aspects of this latter source of uncertainty. However, the first two sources may be more important in many cases. The first one can be addressed by analysis of multiple scenarios. The second one can be addressed by analysis using more than one modeling approach. The third can be addressed using probabilistic methods as described in the Report. Some would argue that the first two could also be addressed by probabilistic methods. The Report indicates that a tiered approach will be applied to probabilistic assessment. The Subcommittee supports this approach and feels that it should receive some more discussion. Appendix A-3 provides a fuller exposition of these issues and includes a substantial bibliography of related literature that will form the basis for a more complete consideration of uncertainty as

the Agency moves forward with actual analyses of residual risk.

Methods for prioritizing data collection include the use of models. For example, sensitivity analysis and probabilistic analysis can be used to help identify key sources of uncertainty, associated with lack of data, upon which risk estimates may be highly dependent. These key sources of uncertainty can be targeted for additional data collection as needed. This approach, in combination with valuation of the cost of collecting the data, provides a systematic method for setting ongoing research agendas. As new data are collected, the need for additional data can be re-evaluated and resources can be retargeted, as needed, to the next most important key sources of uncertainty. The Report is in error in asserting, in several places, that sensitivity and uncertainty analysis can only be performed when there are large amounts of data. Uncertainties are usually of greatest important as data become increasingly scarce, and statistical and other methods exist for dealing with uncertainties in such situations.

In the discussions of control technologies and pollution prevention measures, it is generally important to consider variability and uncertainty in control technology efficacy and cost, in addition to the other sources of variability and uncertainty in exposure and risk assessments. This is particularly true for new MACT technologies where in many cases, the control efficiency will not be known until it is put in place and tested. The probabilistic methods described for exposure and risk assessment are typically general enough for application to technology assessment problems (Frey and Rhodes (1996), Frey et al. (1994), and Frey and Rubin (1998)).

A specific area where data are severely lacking is in the areas of emission rates, emission inventories, and ambient air quality data for the 188 HAPs. Emission measurements for HAPs can be expensive and difficult, which accounts in part for the lack of data. Because HAPs have only recently (compared to CAPs, for example) become the subject of regulatory scrutiny, databases are only now being developed and typically are incomplete. For example, HAPs emissions are often poorly characterized. Data tend to be available only for a small subset of the 188 pollutants, and the quality of such data varies greatly among the 174 source categories. To support both chronic and acute health risk assessments, it is necessary to measure HAPs emissions for long time periods using short sampling times; e.g., years of hourly or daily data. In the absence of such data, many assumptions (judgments) will be needed in order to make estimates of emissions for averaging times that are appropriate to health and ecological risk assessments. The use of judgment is inherent in any risk assessment process and should be recognized and made as transparent as possible. The Report should refer to the Agency's current steps to develop a National Toxics Inventory and discuss how this effort will address the SAB's concerns.

In order to develop emissions estimates for use in risk assessment, it will be necessary to consider not just emission factors or emission measurements at a representative set of facilities for each of the source categories, but also to consider the activity levels of the emission sources within the geographic scope of each assessment in order to develop an emission inventory. By activity, we mean the processes and subprocesses within a facility that gives rise to the emitted HAPs. An inventory is typically conceptualized as the product of an emission factor (e.g., mass

emission of a pollutant per unit of activity associated with the release of the pollutant) and an activity factor (e.g., the number of units of activity), summed over all emission sources. Data on activity factors can be difficult to obtain. For many source categories, activity is highly variable, especially over short averaging times. In addition, because activity data may be difficult to obtain, there is often substantial uncertainty regarding activity levels. Thus, the collection of activity data may become an important priority for improvement of risk assessments. The importance of the collection of emissions and activity data can be assessed via sensitivity and probabilistic analyses, as noted previously.

The development of risk estimates will likely rely heavily on the use of dispersion models. It should be noted that the typically employed Gaussian-based dispersion models are considered to be precise to no better than plus or minus 30 percent and are only appropriate for evaluation of short-range transport (less than 50 kilometers). The preliminary results from the Assessment System for Population Exposure Nationwide (ASPEN) modeling effort suggest that uncertainties may be far greater than plus or minus 30 percent (Rosenbaum and Cohen, 1998). There is a lack of validation of such dispersion models in most cases. Furthermore, it is not likely that the dispersion of all HAPs can be easily or appropriately modeled using Gaussian plume models, due to their chemical reactivity and/or physical characteristics. The comparison of air quality modeling predictions, which will included uncertainties associated with emissions estimates, stack parameters, meteorological scenarios, and the structure of the models themselves, with measured ambient monitoring data, is an important means of providing insight into the precision and accuracy of the models. Thus, efforts should be continued and expanded regarding the collection of ambient HAPs measurements.

Precision refers to lack of unexplained variability in model predictions, whereas accuracy refers to lack of systematic bias in model predictions. The precision and accuracy of dispersion models should be quantified and considered as a source of uncertainty when performing exposure and risk assessments. Results from the ASPEN effort may be useful in this regard, although, as previously noted, these results suggest that the precision of the existing dispersion models is rather poor.

It is also important to develop a sound basis for estimation of background levels. At this time, the estimation of background levels is highly uncertain and perhaps even speculative in many cases. A program of additional measurements should be considered as a means to improve the database and reduce uncertainty regarding estimation of background levels. Since the role of background is a problem that surfaces throughout the Clean Air Act and other environmental legislation, the Report should draw upon the experiences of other programs also.

It is important to define the risk characterization endpoints prior to performing a significant number of analyses. In fact, the definition of endpoints is needed early on in order to help anticipate data collection and research needs in support of the Residual Risk program. For example, the evaluation of various health and ecological endpoints will have implications for the temporal and spatial characteristics of each assessment. For acute endpoints, data based upon

short averaging times (e.g., hourly, daily) will be required for all assessment inputs. For chronic endpoints, data based upon long averaging times will be required. As noted elsewhere, for example, emissions data may in some cases be available for a convenience sample of short averaging times (e.g., daily), but collected only over a short testing program (e.g., only for a few days). In this example, temporal patterns in emissions (e.g., seasonal variations and autocorrelations) would likely not be revealed. Thus, emissions data collected over a short duration for only a limited number of short time periods may not be adequate for supporting acute risk assessments, nor would it be a sound basis for making chronic risk assessments. The geographic scope of assessments also has important implications for data collection. If localized and acute health effects are to be studied, then highly location-specific data may be required. In contrast, if chronic effects that may result from longer range transport are of importance, then "representative" regional or national average data may be sufficient. Variability and uncertainty tends to increase as the averaging time or geographic scope of a study decreases.

Because it is unlikely that all data gaps will be filled prior to the development of residual risk estimates, it is important to consider and employ methods for the quantification of both variability and uncertainty. In fact, an EPA-sponsored Workshop in April, 1998 provides insights on these matters (See Appendix A-3).

In order to more realistically manage the residual risk requirements, it will be necessary for the Agency to prioritize the focus of the assessment effort. Prioritization may be easily accomplished by screening the list of 188 HAPs to identify those that are least active in terms of human and ecological health effects and to focus initially upon those that appear to pose the greatest threats. Similarly, EPA should prioritize the 174 source categories not merely based upon the timing of implementation of MACT standards for those categories, as dictated by Congress, but based upon screening-level assessments of which source categories may pose greater residual risks than others. As new data become available, the screening studies should occasionally be revisited to make sure that no important HAPs and/or source categories are overlooked.

It should be anticipated and stated that there is uncertainty regarding both the MEI and the MIR. It is appropriate to constrain the MIR to be representative of an actual person, rather than a fictitious "porch potato" or resident in the middle of a lake. However, it is also important to consider the 1992 Exposure Assessment Guidelines and include the notion of a high end exposure and a mid-range exposure in assessments beyond the screening stage. These can easily be inferred from the results of probabilistic analyses. Since uncertainties tend to be greatest at the extreme tails of distributions, measures such as MIR are likely to be highly uncertain compared to average population risk characteristics. In fact, the range of uncertainty for the MIR is likely to be very large compared even to the risks associated with high-end exposures; e.g., around the 90th percentile. It is not realistic to expect any method to be able to make a precise prediction of the MIR, and this should be clearly stated in the Report.

Because uncertainties in risk assessments are typically large (Talcott, 1992), there is a special challenge for the evaluation of unintended consequences. When comparing two risks, it can be difficult or impossible to determine which one is really higher, because both may be uncertain by orders-of-magnitude and have overlapping uncertainty ranges. Probabilistic methods, if properly employed, can help provide an indication of the likelihood that one risk is really higher (or lower) than another risk (Appendix N-2 in NRC, 1994). However, it should be expected that the results of such assessments may not be conclusive in many cases.

3.2.2 Ecology

3.2.2.1 General Comments

As noted above, the Subcommittee endorses the general RA/RM strategy laid out in the Report for addressing health and ecological residual risks and compliments the Agency on these initial efforts. While the practical applications of the RA/RM approach are not as fully developed with respect to ecology as it is to health, the Agency has made significant strides in the past few years to provide a sound technical basis for such assessments in the realm of ecology. The Agency needs to continue to grow in its appreciation of the role of ecology in its corporate mission and to place appropriately increased priority on examining the impacts of stressors on the ecology. In contrast, some Subcommittee members detected an unfortunate, apologetic tone in the Report's description of ecological risk assessment. The Subcommittee believes that the Agency is at the forefront in the development of ecological risk assessment methods. The challenge will come in applying these new tools to residual risks assessments in ways that are generally new to the field.

The field of ecological risk assessment has deep roots in the aquatic sciences. With the exception of the assessments performed on pesticides, the strength of data and experience is in assessing the risk of chemicals to aquatic systems. Even in the performance of risk assessments for terrestrial systems, we often rely on extrapolation from existing data for aquatic organisms; e.g, Ambient Water Quality Criteria. The ability to assess non-pesticide chemical risks in terrestrial systems is advancing, but mostly in assessing soil-bound sources of contamination. There is a body of knowledge on the atmospheric fate, transport, and environmental impacts of a few key persistent organic pollutants, mostly pesticides and halogenated chemicals. However, extrapolation from our knowledge of these chemicals to the 17 HAPs classes should be done with caution. There are many ecologists and ecosystem mangers who have had significant experience in analyzing management needs at regional levels whom the Agency can consult in framing the Residual Risk question.

The Report, as it stands, is too general to be very helpful in a practical sense. At certain points, it bears a resemblance to a guidance document. However, its discussion of issues is couched in terms that are noncommittal, vague, and/or elementary. Consequently, it is difficult to ascertain what will be done when the Report is applied to specific cases. This is more evidence in the section on ecological risk assessment than it is in the human health section.

Specifically, the ecology sections of the Report are often times couched in very simplistic terminology and this may be problematic. There are some portions that simply do not reflect accurately current ecological theory and practice. The issue of how HAPs are treated in the lowest level tier is an important illustration of the weakness. The Report offers two criteria for identifying HAPs that should receive special attention: a) potential for bioaccumulation and b) lifetime. These are important, but by themselves these two criteria are insufficient. Two examples illustrate the point. First, ozone is one of the most significant regional pollutants affecting ecological resources and human health. However, its residence time in the atmosphere is minutes to hours, and its bioaccumulation potential is zero. Thus, ozone would not be identified in the initial screening exercise for ecological effects; i.e., a false negative. Second, consider chlorofluorocarbons (CFCs), which have a long residence time, but their residence time is in the stratosphere (where their breakdown products scavenge ozone and thereby increase ground level ultraviolet radiation) rather than in the earth's crust or the biosphere where they could affect life, for which CFCs are generally non-toxic. Again, an initial screen could possibly generate a false negative for the pollutant. In short, the Agency should develop a more robust approach to the first tier of screening criteria for ecology and ensure that the criteria strike an appropriate balance between the probability of a significant number of false negatives or false positives. The two proposed criteria are a reasonable starting point, but they need to be amplified. Other criteria might include inherent toxicity, significant contribution to criteria pollutant levels, potential reaction products, and partitioning in the environment (e.g., octanol-water partition coefficient (K_{ow}), deposition to canopies).

As another example, it is not clear that the Report adequately appreciates the issue of residence times in the environment and the scale of ecological analysis. For example, if the chemical has a residence time of a month or more, then the distribution of the chemical will approach hemispheric proportions. Longer residence times in the atmosphere will lead to global distributions. Will the scale of the ecological and human health risk assessment be scaled according to the atmospheric residence time? This approach raises the issue of our effect on others around the globe and their effect on us. For example, consider the case of mercury. Given that the residence time of mercury in the atmosphere is one year, mercury is almost by definition a global problem. Long residence times again raises the question of the background concentrations. Therefore, policy makers need to consider the degree to which reduction of US emissions of mercury will reduce US risks from mercury in the environment.

The Report applies a hierarchical theory of ecology. That is, the Report places an emphasis on the "scaling up" approach in which data collected at the molecular and individual level are translated to higher scales (population, community, ecosystem). This approach assumes, without discussion, that analyses at one scale of hierarchy are directly applicable to scales higher up (or lower down). The Report should accompany any discussion of trans-scale applicability of analyses with references to the scientific literature that support that perspective. If such literature is not available, then this approach is a default assumption analogous to those in the health risk process and should be identified as such.

Similarly, the only technique that is presented to address broader scales in ecology (e.g., landscape, watershed, and ecosystem) is the scaling up approach. While the scaling up approach should certainly be discussed, there are other methods that can be used independently or used in conjunction with others. Examples include modeling, geographic information systems, ecological epidemiology, remote sensing, and landscape ecology.

The Report argues that society is not concerned about mortality of individuals among species, except in the case of humans. In fact, there are stakeholders who would very much disagree with such a proposition. Notable examples in which the plight of individuals or small groups of individuals has prompted considerable public concern include members of rare and endangered populations (e.g., panthers in Florida), animals in highly valued ecosystems (e.g., wolves in Yellowstone National Park), and animals that become the focus of media attention (e.g., marine mammals beached within easy reach of TV cameras).

3.2.2.2 Charge Element 2

The document is replete with qualitative statements about ecological RA and how it be done. In general, quantitative efforts are downplayed, and there is no significant discussion of which quantitative data will be gathered, analyzed, and used in the risk assessment process. As in any effort, providing a boilerplate framework does not provide enough guidance on the quantitative aspects of the risk assessment strategy. In short, the quantitative rigor of risk assessment and management for ecosystem risk should be given more visibility. The issue is not one of presenting the quantitative aspect in this report but rather, making sure that the audience appreciates the degree to which quantitative analyses will be performed.

The Agency is building upon a solid set of accomplishments that have put ecological RA on a sound footing. However, the Agency is breaking new ground when it is addressing the risk that atmospheric releases pose to ecological resources. An indication of this comparatively new ground is the fact that of the 17 cases studies provided by the Agency during the development of the ecological RA guidelines, only two of them -- one on "acid precipitation" and the other on "ozone" address regional or landscape scale impacts of contaminants from atmospheric sources. In both cases, there were clear linkages between the assessment endpoints, the environmental impact, and the stressor of concern. The Report does not provide a similar level of sophistication and confidence for HAPs with regards to the definition of the problem, the management goals, and the assessment endpoint(s).

A major barrier to the successful application of the Agency's ecological RA methods to the Residual Risk case is the paucity of information provided with regard to management context and problem formulation. These matters are addressed in only the most general fashion. There is no real attempt to clearly define "What is being protected?" or "What constitutes an adverse environmental effect?" Without such important details, the analysis plan that follows is too vague to evaluate properly.

The Agency's approach of using no effect concentrations (NOECs) as levels of concern, coupled with the use of hazard indices (HIs) calculated for effects to sensitive individuals, results in an ecological RA that is designed to protect the most sensitive individual. This is not what Congress seems to intend by the language used in Section 112(1)(7) of the CAA where the concern is couched in terms of "...adverse impacts on populations of endangered or threatened species or significant degradation of environmental quality over broad areas." If the Agency's goal of their risk management is to protect all individual organisms, then it should state that goal clearly. However, such a stringent management goal seems misstated or misguided. In fact, the Agency's proposed reliance on the simplistic "quotient" or HI approach raised some concern. While an argument could be made for its use as a crude screening tool, there are significant problems with the approach in higher level decision making steps. A major limitation of the quotient approach is that it provides a point estimate of the risk and is clearly a one dimensional model which relies on concentration (Suter, 1993). Such a model, yields very limited information for a risk management decision. This limitation is clearly spelled out in the Agency's ecological RA guidelines (Section 5.1.3 in USEPA, 1998). More fundamentally, there is an area of debate on whether risk calculations should be expressed as point estimates (cf., the HI approach) or as probabilities and ranges. The use of point vs. probabilistic values is an underlying philosophical issue that should be dealt with more generally in the Report.

The current statement in the Report on "assessment endpoint" (Sec. 5.4 p. 114) is too vague to be useful in establishing measurement endpoints and ultimately assessing the success of risk reduction strategies. The Report should provide a clear definition of an assessment endpoint or recognize that it is not possible to do so, except in the context of the assessment for a specific source category.

The Report overlooks the value of epidemiologic or field data in demonstrating the presence or absence of a cause-effect relationship that can provide a basis for prioritization or recognizing the efficacy of any management strategy. Without a clear link between the stressor (HAPs) and the effects, it is not clear how the Agency will ever design a realistic risk assessment strategy or test the success of any risk reduction actions triggered by their risk assessments. Some field studies in regions that are data rich in monitoring or other types of resource data could lead to the development of an effective residual risk assessment and management program.

Ideally, the HI would replaced by some fundamental knowledge of effects at the molecular level, thereby obviating the use of the HI as a means of addressing effects of mixtures. And yet, the field of ecology is a long way from having such knowledge for most pollutants. Accordingly, what default methodology will be employed? It is unrealistic to default to a molecular mechanism of toxicity as the means of addressing mixtures (or individual chemicals) for ecological risk assessment. As an alternative, consider that for most pollutants at chronic levels in ecosystems, the adverse effects are largely mediated through some ecologically/physiologically significant process that governs fitness; e.g., photosynthesis in plants, respiration in animals, or reproduction. Therefore, the Agency should explore formulating its ecological risk assessments by considering how the chemical (or mixture) affects critical processes governing fitness.

To the degree that the Report refers to the far greater in-depth analysis of risk that will take place in the higher tiers, the document is short on specifics. In fact, there is no road map that details the quantitative nature of this effort. Brave statements that the higher tiers are more quantitative and accurate are not supported by substantive discussion. The Agency should develop additional guidance of what will occur in the higher tiers, especially as it relates to quantitative assessments and uncertainty analysis. Such guidance would, of course, benefit from critical peer review.

Presumably, higher tier assessments will incorporate additional factors, such as social and economic concerns, some aspects of which can be subject to technical analysis. The Report is silent on whether, how, and by whom these analyses would be conducted. This absence is an important gap in the description of how risk assessment will be used in the risk management process in the Residual Risk program.

Arguably, the role of conceptual models is overplayed in the Report. While such models have value, particularly in risk communication, their role is largely qualitative, as described, and is more limited in the risk assessment, per se. By contrast, the use of quantitative/simulation models to investigate the behavior of ecological systems is underplayed. While computer models are acknowledged in the transport, transformation, and fate sections, the fact is that the ecological science has come quite far in the development and use of simulation models to address effects on ecological systems. It is this aspect of modeling that should play a more prominent role in residual risk assessment. For example, such models are particularly appropriate for analyses that at the watershed and regional level.

3.2.2.3 Charge Elements 3 & 4

The Report briefly addresses (in section 3.3.2) the various types of data that might be used in an ecological RA. For effects characterization, the Report lists 1) field studies, 2) microcosm studies, 3) laboratory studies, and 4) structure-activity relationships. However, there is no indication of the availability of such data for the HAPs that the Agency will use. Later in the text, there is a general discussion about what is needed for ecological exposure characterization. But, again, specifics are lacking. There is even an intriguing allusion to some new approach being developed by OAQPS, but no details are provided. The final section of the Report (USEPA, 1998, Section 5.4, pp 112- 122) contains further reference to types of data which may be required. However, without the clarify of a process map or analysis flowchart, the reader is left guessing about data will be needed, its source(s), and its use. While the Report contains a number of sources for obtaining data and methods for risk assessment, several prominent ones are missing. For example, the use of quantitative models of the ecosystem, the emerging field of ecological economics, and the field of ecological epidemiology are not adequate presented.

The fact is that there is a real lack of data for effects endpoints, especially for plants, birds and wildlife. This constraint should be stated. If the Agency intends to fill these data gaps by either extrapolation or projection modeling, they should provide a clear definition of the models

they will rely upon or will need to develop. The Report should contain some sense of the time necessary and available to develop and/or validate the needed models.

There is a paucity of good benchmarks for environmental effects. There are no comparable IRIS or HEAST databases for ecological effect endpoints which provide RfD values. Benchmarks are more readily available for aquatic organisms but are also being developed for some terrestrial organisms. Little, if any, peer review has been performed for any of these benchmarks. Generally, the benchmarks were developed for a specific risk management purpose which may or may not overlap with the management of post-MACT residual risks. The technical assumptions, defaults and methodologies used to calculate any set of benchmarks are often forgotten and the numbers gain a life of their own. The report should reflect this current state-of-the-art for using ecological benchmarks. The Agency could address this issue by developing its own benchmark methodology and set of numbers that align with the risk assessment of the residual risk of HAPs.

In short, the Report needs to clearly state (and illustrate with examples, as appropriate) what data the Agency is planning to use, where they will come from, and how they will be used. Further, if there are new approaches under development, they should be described. Such a clear description of the situation would instruct Congress about the challenges that the Agency faces in providing realistic estimates of risk that will form the basis for decisions on risk reductions.

3.3 Charge Element 5

5. Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program?

The entire process of risk assessment is oriented toward managing the risk. The discussion in the Report on <u>how</u> risk mangers will utilize the scientific and non-scientific information in making their decisions is quite abbreviated. There is no discussion of the approach that will be used by the manager in deciding what to do and how to proceed. Risk management is a critical activity, and its technical process and content steps should be as clearly addressed as are the steps involved in risk assessment.

A major part of the risk assessment methodology will involve the use of models to estimate various results that are needed to calculate the risk. However, there is a major problem in the failure to validate models before employing them in the regulatory setting. The report indicates that an adequate validation of Human Exposure Model (HEM) has not been performed. This lack of validation of earlier models leads to question as to what extent new models like Total Risk Integration Model (TRIM) can and will be validated. Particularly given the short time available, it is not clear that it will be possible to even complete the development and initial testing of TRIM. It appears to be a common problem that Agency models are inadequately tested and validated before they are applied to regulatory decisions. There needs to be adequate testing and validation of any model before applying it to actual problem solving. It seems very unlikely that

the Agency can develop, test, and validate any new model within the time frame available. This raises considerable uncertainties as to the validity of the results that will then be important components of the residual risk analyses

The Report is also comparatively silent on how stakeholders will be involved in the risk assessment and management process. There needs to be some language as to how stakeholders will be identified, represented, and involved. These approaches, strongly espoused in the S&J and CRARM reports, should be based on sound social science principles.

In general, the Reference List is inadequate. For instance, over two-thirds of the references listed are Agency or CAA required or commissioned studies. A more robust Reference List with contemporary, peer- reviewed articles would give the general reader more confidence and the informed reader greater access to further information. Citations to specific examples would be especially helpful; cf., experience in working with the risk assessment of HAPs at a regional scale, such as the Agency's experience with acid rain and ozone should be included. Even if it is not possible to provide specific examples, a reading list of additional related peer-reviewed reports of risk assessments should be provided. This list would provide the reader with an indication that risk assessments can be done and some idea of the type of assessments that have been performed and accepted in the past. A section that references existing State air toxics programs would greatly enhance the Report. For example, the California State Air Toxics Program is one of the most comprehensive in the country, and a more thorough overview of this specific State air toxics program should be referenced, if not discussed in an appendix to the Report. The use of a broader literature is particularly critical given several questions that have been raised about the lack of acceptance in the scientific community of some of the approaches advocated in the Report. Thus, there should be some means of not only drawing upon a wider literature, but also developing a process to ascertain the prevailing opinions of the scientific community (if not consensus) on many of the issues raised in the Report.

The Report would benefit from a generic process map that provides a representation of how the Agency intends to prioritize individual HAPs for analysis, how tiers will function, and how data needs will change between screening level assessments and more definitive risk assessment efforts. Much of this information already exists in Sec. 5.4 and Exhibit 18 (page 120). Some reorganization of the information and a more schematic presentation of the information would be helpful.

The Report is quite comprehensive in its scope. However, sometimes the purpose of the Report gets lost. It is not always clear whether the Report is a general review of current risk assessment methods or a more sharply focused discussion of RA methods as applied to the question of residual risk. The distinction, if any, between RA in general and RA for the purposes of Residual Risk should be made clear. The strategy presented at the end of the report (Section 5) should incorporate more of the elements mentioned throughout the text as they relate to Residual Risk program.

The Report should be more explicit about what it is -- and is not -- providing. For example, the Report cannot solve all of the outstanding risk assessment problems and must be selective, focusing on what is most relevant to residual risk. Instead of providing such a focus, however, Section 3 primarily reviews and critiques most of the methods that exist. The reader does not know what decisions have been made for implementing the Residual Risk program.

As noted above, the Subcommittee applauds the use of an iterative screening technique as an initial step for the analysis. This approach is a useful means of simplifying the process and seems to have gained wide support. The iterative screening technique would be strengthened by greater specification of the procedure, what it depends on, and how it is applied specifically to Residual Risk.

4.0 Conclusion

The Agency's Report is a useful strategic document that will help guide the Agency as it moves ahead with the Residual Risk program. However, the Subcommittee recommends that the Agency be more candid with Congress and the public about what can be accomplished with existing limitations in data, models, methods, time, and resources. The Subcommittee has pointed out many areas that will require more thought, more documentation, and more articulation when the program is actually implemented.

REFERENCES

Burke, T., "Public Health and Environmental Protection", presentation at the Annual meeting of the Society for Risk Analysis meeting in Washington, DC, 1998.

Csanady, G.A., F.P. Guengerich, and J.A. Bond, "Comparison of the Biotransformation of 1,3-Butadiene and its Metabolite, Butadiene Monoepoxide, by Hepatic and Pulmonary Tissues from Humans, Rats and Mice", Carcinogenesis 13, 1143, 1992.
[Ref on p. 13]

Commission on Risk Assessment Risk Management (CRARM), "Framework for Environmental Health Risk Management", Final Report, Volume 1; "Risk Assessment and Risk Management in Regulatory Decision-making", Final report, Volume 2, 1997.

Frey, H.C., E.S. Rubin, and U.M. Diwekar, "Modeling Uncertainties in Advanced Technologies: Application to a Coal Gasification System with Hot Gas Cleanup", Energy 19(4): 449-463, 1994.

Frey, H.C. and D.S. Rhodes, "Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example", *Human and Ecological Risk Assessment: An International Journal*, 2(4), December, 1996.

Frey, H.C. and E.S. Rubin, "Uncertainty Evaluation in Capital Cost Projects", in Encyclopedia for Chemical Processing and Design, Vol. 59, J.J. Mcketta, ed. Marcel Dekker, New York, 1997, pp. 480-494.

Hawes, Jonathan, Dewitt John, and Richard Minard, Jr., "Resolving the Paradox of Environmental protection", *Iss. in Sci. and Tech*, 14(4), 57-64, Summer, 1998

National Research Council, Risk Assessment in the Federal Government: Managing the Process, National Academy Press, Washington, DC, 1983.
[Ref p. 7]

National Research Council, "A Paradigm for Ecological Risk Assessment", in Issues in Risk Assessment, Committee on Risk Assessment Methodology, National Academy Press, Washington, DC, 1993.

National Research Council, Science and Judgment in Risk Assessment, National Academy Press, Washington, DC, 1994.

National Research Council, Understanding Risk, National Academy Press, Washington, DC, 1996.

Paustenbach, Dennis J., The Risk Assessment of Environmental and Human Health Hazards: A Textbook of Case Studies, John Wiley and Sons, New York, 1989

Presidential/Congressional Commission on Risk Assessment and Risk Management (CRARM) Report, Washington, DC, 1997.CRARM report

Omenn, G.S., "Putting environmental problems into public health context", *Public Health Reports*, 111:514-516, 1996.

Risk Policy Report, "Critics Argue EPA Failing to Upgrade Key Database", page 23, July 17, 1998.

Rosenbaum, A. and J. Cohen, "Extrapolation of Uncertainty of ASPEN Results (Revised). Draft Technical Memorandum" prepared by ICF Kaiser for D. Axelrad and T. Woodruff, Office of Policy, Planning, and Evaluation, USEPA, Washington, DC, May 13, 1998.

Science Advisory Board, "Interim Methods for Development of Inhalation Reference Concentrations", EPA-SAB-EHC-91-008, 1991.

Science Advisory Board, "Clean Air Act Section 812 Retrospective Study entitled 'The Benefits and Costs of the Clean Air Act, 1970 to 1990'", EPA-SAB-COUNCIL-LTR-97-001, October 23, 1996.
[Ref on p. 4, 11]

Science Advisory Board, "Draft Retrospective Study Report to Congress Entitled 'The Benefits and Costs of the Clean Air Act, 1970 to 1990'", EPA-SAB-COUNCIL-LTR-97-008 , July 8, 1997
[Ref on p. 4, 11]

Science Advisory Board, Environmental Health Committee report on Inhalation Reference Concentrations, in preparation, August, 1998a.

Science Advisory Board, Environmental Health Committee on Acute Reference Exposure Methodology, in preparation, August 1998b.

Science Advisory Board, "Integrated Environmental Decisionmaking", Report from the Integrated Risk Project, pending, 1998c.
[Ref on p. 15]

Seaton, M.J., Schlosser, P.M., Bond, J.A., and Medinsky, M.A. (1994) "Benzene Metabolism by Human Liver Microsomes in Relation to Cytochrome P450 2E1 Activity", Carcinogenesis 15, 1799-1806, 1994.
[Ref on p. 13]

Seaton, M.J., P.M. Schlosser, J.A. Bond, and M.A., "In Vitro Conjugation of Benzene Metabolites by Human Liver: Potential Influence of Interindividual Variability on Benzene Toxicity", Carcinogenesis 16, pp. 1519-1527, 1995.
[Ref on p. 13]

Suter, G.W. II, Ecological Risk Assessment, Lewis Publishers, Chelsea, MI, 1993
[Ref on p. 24]

Talcott, F., "How Certain is That Environmental Risk Estimate", Resources, Resources for the Future, Washington, DC, no. 107, pp 10-15, Spring, 1992.

USEPA, "Alpha-2u-globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat", Office of Research and Development, Risk Assessment Forum, PB92-143668/AS, EPA/625/3-91/019F, September 1991.
[Ref on p. 7]

USEPA, "Hazardous Air Pollutants: Profiles of Noncancer Toxicity from Inhalation Exposures, Office of Research and Development, Washington, DC, EPA/600/R-93/142 1993.
[Ref circa p. 18]

USEPA, "Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry", Office of Research and Development, Washington, DC, EPA/600/8-90/066F, October 1994.
[Ref on p. 12]

USEPA, "Strategic Plan for the Office of Research and Development," Washington, DC: EPA/600/R-96/059, p. 3, May, 1996. Also in "Update to ORD's Strategic Plan", EPA/600/R-97/015, April 1997.

USEPA, Final Report to Congress on Benefits and Costs of the Clean Air Act, 1970 to 1990, Washington DC, EPA 410-R-97-002, 1997.

USEPA, "Assessment of Thyroid Follicular Cell Tumors", Office of Research and Development, Risk Assessment Forum, EPA/630/R-97/002, Washington, DC, IN PRESS.
[Ref on p. 7]

USEPA, "Methods for exposure-response analysis for acute inhalation exposure to chemicals. Development of the acute reference exposure. External Review Draft", Office of Research and Development, Washington, DC, EPA/600/R-98/051, April 1998.
[Ref on p. 12]

USEPA, "Draft Residual Risk Report to Congress", Office of Air Quality Planning and Standards, April 14, 1998.

USEPA, "Guidelines for Ecological Risk Assessment", Office of Research and Development, Risk Assessment Forum, 63 Federal Register (93)26846-26924, 14 May 1998.

Zimmerman, R., "Governmental Management of Chemical Risk," Lewis Publishers (and CRC Press), Chelsea, MI, pp. 70-71, 1990.

APPENDIX A

WRITTEN COMMENTS OF SUBCOMMITTEE MEMBERS

Each member of the Subcommittee prepared written comments on the draft Residual Risk Report to Congress, dated April 14, 1998. These comments were shared at the meeting with the other Subcommittee members, the Agency, and the public. After the meeting, some Subcommittee members modified and/or added to their prepared comments and resubmitted them for circulation to the Subcommittee, the Agency, and the public.

This Appendix contains the final written comments from each of the Subcommittee members. These comments are included in this SAB Report so that the Agency and the public can a) benefit from the specific comments and b) appreciate the range of views represented on the Subcommittee. While all of these comments are commended to the Agency for their careful consideration, unless a comment is addressed explicitly in the body of this SAB Report, the comment should not be represented as the collective view of the Subcommittee.

Comments follow from the following individuals in the following order:

Appendix A #1 -- Dr. Gregory Biddinger
A #2 -- Dr. Thomas Burke
A #3 -- Dr. H. Christopher Frey
A #4 -- Mr. Thomas Gentile
A #5 -- Dr. Philip Hopke, Subcommittee Chair
A #6 -- Dr. Michele Medinsky
A #7 -- Dr. Warner North
A #8 -- Dr. Gilbert Omenn
A #9 -- Dr. George E. Taylor, Jr.
A#10 -- Dr. Rae Zimmerman

APPENDIX A-1

Comments on First Draft
SAB Report on USEPA
Residual Risk Report to Congress
Greg Biddinger
August 28,1998

In the following comments you will see a number of common themes to those presented by Dr. H. Christopher Frey in his email review (dated August 27,1998). I n particular I believe that the subcommittee made many strong recommendations which carried the expectation that they need to be included in the next draft in order for the committee to see this as a good report which will be valuable to respond to the Congress' questions in 112(f).

Therefore I strongly support the idea of using bulleted recommendations both in the executive portions and under each of the charges in the body of the report.

The way these recommendations occur now in the report is a soft presentation of some rather strong suggestions. We should make it easy for the agency to understand our recommendations by using a format that allows them to see the big issues raised by the Subcommittee. An example of such a format was the Science and Judgement executive summary. That was a very complicated text and yet without spending a massive amount of energy you got the message about what they thought was needed. As I go down the list I will try to identify the recommendations that I think were emphasized especially for ecorisk.

TRANSMITTAL LETTER

General Comments.

1. **We need to be more careful with the use of qualifiers in the general conclusions in the transmittal letter and need to be clearer on our recommendations and expectations.**

1. Line 37 The sentence " In short, the SAB found the Report to be a good document, but one that could be strengthened in a number of important places, as identified in this report"; should read as follows:

In short, the SAB found that the report to be a reasonable first draft but one that requires strengthening in a number of important places. Aspects requiring improvement are highlighted throughout the report in the form of bulleted recommendations.

The use of the word good in the first sentence could to easily be translated by management as "good enough" If that were the case then there would not be more than 20 pages of observations and recommendations which followed.

The rest of the transmittal letter seems to summarize many of the issues with the correct emphasis. I believe by changing this lead in sentence we will not give the mistaken impression the report only need a minor polishing.

I agree with Chris Frey's recommendation for use of bullets in transmittal letter to emphasize the subcommittee's overall impressions. (reprinted below with minor suggested changes which are underlined)

• *The Report gives a misleading impression that more can be delivered than is scientifically justifiable, given the data gaps and limited resources (e.g., time, funding) for conducting the residual risk assessments. The committee recommends that the Report more carefully convey the limitations of the data, models, and methods that are described or that would be needed to carry out the residual risk assessment activities.*

• *The Report should contain or cite specific examples to clarify what some of the bold, but vague, language is intended to convey. For example, a frank and clear discussion of: (a) current limitations in available methods<u>(e.g. ecological risks at regional ecosystem levels)</u> and data (e.g., emissions, IRIS, HEAST); (b) methods for reducing data gaps (e.g., the promise of uncertainty analysis to value-rank data gaps); and (c) priorities for research and management action should be provided.*

• *The Residual Risk program could evolve into a "paralysis by analysis" activity without an appropriate and well-supported screening approach to prioritize assessments among the 188 pollutants and 174 source categories. It is important that EPA avoid screening methods that generate a large number of "false positives", while at the same time the Agency must avoid excessive attempts to resolve all of the nuances of the complex risk assessment issues for all HAPs and all sources. The Agency needs to carefully prioritize its assessments and husband its resources, lest the program evolve into a wide, but shallow, program that fails to adequately quantify and target residual risks or into a program that fails to address a sufficient number of pollutants and sources due to over-analysis of just a few cases.*

Unlike Chris, I support the inclusion of recommendations for a clarification of connections to management context and the value of examples which would help in this vein. That bullet could be something like

The report would be improved if the value and ultimate use of the risk assessment were clarified by outlining how the results will be applied to make risk management

decisions. Example should be used through out the report to illustrate both the soundness of the science used in the residual risk assessment and the risk management context in which it will be used.

THE ABSTRACT

Editorial changes

Line 16 , page ii delete " For example" . The phrase seems to beg a proceeding sentence and by leaving it out the sentence reads fine.

Line 19, page ii change "which could what" to ***"that could clarify"***

Line 24 , page ii change " should no attempt" to ***Should not attempt***"

General Comments

Suggest repeating the bullets from the transmittal letter or some variation of that theme.

1.0 EXECUTIVE SUMMARY

Editorial Comments

Page 2 line 39 Change "would" to "should"

General Comments

1. **Suggest repeating the bullets from the transmittal letter and to incorporate a format similar to that used in Science and judgement executive summary. Where we provide the bulleted recommendations separately under each charge** These recommendations should be brought forward from the sections in the back and then used as headers in the sections in the back. I have tried below to identify those recommendations but others should double check to make sure I am not missing any of their key recommendations.

2. Page 1, lines 42-43. Delete the sentence " In general, the Agency has generated a good report to congress that meets the requirements of section 112(f)(1) of the Clean Air Act as amended and replace with:

"In short, the SAB found that the report to be a reasonable first draft but one that requires strengthening in a number of important places."

This will provide consistency with the transmittal letter and the abstract. As I said before it is not appropriate to call the report good as it might be interpreted as good enough. Also the committee never discussed if this report satisfied the requirements of section 112(f)(1). That was not part of our explicit charge. If this is a needed assessment of our committee then we should reconvene by conference call to assess this point. My current position is that it does not adequately get the job done but a committee level discussion might convince me otherwise. My real concern is that the reader will tie leave with the impression that this SAB subcommittee thinks this report gets the job done that congress wanted and that it is good enough as it stands. I don't feel that way and would be surprised based on the discussions during the 8/3/ 98 meeting if others feel it hits this target.

3. Page 2 Line 9-10. Delete the sentence " Even in the face of less than ideal information and tools, however, the Agency should be able to generate useful, credible risk assessments.

The terms useful and credible are value laden terms which beg criteria. I don't remember us ever coming to this conclusion during our discussions. The statements may be marginally true for Public health depending on your criteria for useful and credible, but they certainly are not true for ecological risk assessments to assess widespread adverse effects to populations. We have great difficulty doing that for chemical stressors where we have lots of data and knowledge of the stressor-effects relationship and modes of action (e.g. acid rain). We are not prepared to do this for the 188 HAP's at this point. In 10 years with concerted efforts 1) to build and validate models and 2) to develop the need chemical specific data on fate and effects, we may be able to say this sentence is true. I suggest deleting it because I don't have an alternative.

4. Page 3 Lines 1-5 . Delete this paragraph and replace with the one as follows:

In summary, if the Agency were to adequately address the proceeding recommendations then this report will provide congress with a useful report. Congress will be able to assess the agencies ability to evaluate the residual risks after the implementation of 112(d) of the Clean Air Act and to take action as necessary to provide the time and resources the Agency needs to accomplish the task. Congress gave them.

Recommended bullets by section for use in executive summary and as headers in the appropriate sections

Introduction:

- The Report gives a misleading impression that more can be delivered than is scientifically justifiable, given the data gaps and limited resources (e.g., time, funding) for conducting the residual risk assessments. The committee recommends that the Report more carefully

convey the limitations of the data, models, and methods that are described or that would be needed to carry out the residual risk assessment activities.

- The Report should contain or cite specific examples to clarify what some of the bold, but vague, language is intended to convey. For example, a frank and clear discussion of: (a) current limitations in available methods (e.g. ecological risks at regional ecosystem levels) and data (e.g., emissions, IRIS, HEAST); (b) methods for reducing data gaps (e.g., the promise of uncertainty analysis to value-rank data gaps); and © priorities for research and management action should be provided.

- The Residual Risk program could evolve into a "paralysis by analysis" activity without an appropriate and well-supported screening approach to prioritize assessments among the 188 pollutants and 174 source categories. It is important that EPA avoid screening methods that generate a large number of "false positives", while at the same time the Agency must avoid excessive attempts to resolve all of the nuances of the complex risk assessment issues for all HAPs and all sources. The Agency needs to carefully prioritize its assessments and husband its resources, lest the program evolve into a wide, but shallow, program that fails to adequately quantify and target residual risks or into a program that fails to address a sufficient number of pollutants and sources due to over-analysis of just a few cases.

- The report would be improved if the value and ultimate use of the risk assessment were clarified by outlining how the results will be applied to make risk management decisions. Example should be used through out the report to illustrate both the soundness of the science used in the residual risk assessment and the risk management context in which it will be used.

Charge 1.

- Explain to congress the large uncertainties and judgmental basis for cancer risk numbers in default assumptions, such as low-dose linearity and the importance of these issues in risk assessment

- Acknowledge the uncertainty regarding whether the dose-response relationship for carcinogens (and some non-carcinogens) at low doses is linear or nonlinear.

- The Agency needs to follow the recommendations in S&J to improve criteria for defaults and for the departure from default assumptions.

- Case studies should be included to as very useful devices for demonstrating how an iterative, tiered process actually works.

- The substantial limitations of the IRIS data as outlined in S&J should be reviewed in the report to congress and recommendations for improvement provided.

- Provide emphasis on setting priorities research and further data collection as an output from the iterative , tiered approach.

- The Agency should convene a workshop to evaluate the degree to which the recommendations in S&J are also applicable for ecological risk assessment and make recommendations for improving the methodology

- The risk management context in which risk assessment s will be used to make decisions should be more explicitly described.

- As a confirmation of the how the Agency considered the CRARM recommendations, the Agency should list them all in a comparative table in Section 5.3.5

Charge 2. Health

- The Agency needs to provide a clearer definition of the difficulties involved in assessing residual human health risks as a result of exposure to mixtures of chemicals from multiple pathways.

- Communicating to Congress the limits of our knowledge and risk assessment tools is essential in order to prevent the misconception that we know more than we do. These limitations include 1) many methods are in a developmental stage; 2) rudimentary knowledge of complex toxicological interactions of mixtures at low doses and 3) the incomplete nature of databases for validating models and assessing toxicological effects.

- The report should clarify that the need for risk assessment for acute non-cancer risks is related to the averaging times dictated in various regulations (e.g. annual averages).

- The Agency should resolve the ambiguity about how it plans to use the Benchmark Dose (BMD) as an alternative to the NOAEL to determine a dose without appreciable effect, based on experimental data.

- The Agency should clearly acknowledge that the use of surface area as a scaling factor is a default assumption used in the absence of chemical-specific knowledge about metabolic activation.

- The Agency should acknowledge that the Agency's methodology for assessment of risks from chemical mixtures is currently under review and changes are possible.

- The report should make it clear that the combining of effects from different chemicals is a technical policy, explicitly designed to generate an excess estimate of risk for screening purposes, and is not based on science and as far as we can tell is not consistent with the Agency's own guidelines for assessing the risks from chemical mixtures.

- The Agency should seriously compare its philosophy and methods for protecting public health with the approach evolved by the Public Health Community (PHC). That is to say they should focus on reducing the incidence of the stressor regardless of the source.

- The Agency should build its residual risk program as a natural extension of the MACT program, Benefiting from the experience gained from the efforts already underway.

- The report should indicate that the Agency will investigate more closely those state and local air toxics programs that are already grappling with residual-risk related problems.

General Comments. Ecology

- The Agency should stress in its report to congress that there is limited experience with performing ecological risk assessment on atmospheric sources of chemicals over regional environmental systems. In general, the Agency as a whole seems to be working at the cutting edge of ecological risk assessment. The Residual Risk program will challenge the Agency's developing abilities.

- For regional ecological risk assessments the Agency may want look to the experience of ecosystem managers.

- The discussions of ecology in the report are very vague, elementary and a bit simplistic they should be improved and made less noncommittal.

- The two screening criteria of bioaccumulation and lifetime may not be adequate to assess the effects of pollution at a regional level.

- The report does not adequately appreciate the issue of residence time in the environment and the scale of ecological analysis.

- The Agency adopts a hierarchial theory of ecology and emphasizes a "scaling up" approach from toxicology at the individual level to effects at the ecosystem level. The assumption is not support by literature related to trans-scale applicability of such data. The literature basis needs to be provided or at least recognized as a default assumption not yet supported by based in scientific study.

- The Agency should not assume that society is not concerned with loss of individuals. In some cases threatened and endangered individuals or large vertebrates (e.g. Florida panther) may drive the assessment.

Charge 2. Ecology

- The report is overly qualitative and the quantitative rigor of the assessment and the management of ecosystem risk should be given more visibility.

- The Agency does an inadequate job of providing a sophisticated definition of the potential environmental problems associated with exposures to HAP's. The possible range of risks, management goals and their potential assessment endpoints needs to be clearly defined and discussed more fully.

- The Agency needs to clearly define: 1) what is being protected and 2) what constitutes an adverse ecological effect from an exposure to HAP's. If the Agency's goal is to protect each member of any wildlife population it should state that goal clearly.

- The Agency's use of Hazard indices (HI's) based on no effect concentrations (NOEC) to sensitive individuals in the population results in an ecological risk assessment designed to protect the most sensitive individual. This is in direct conflict with Section 112(1)(7) of the CAA that focuses the assessment on adverse impacts to populations. The Agency should address this conflict and state why it has not relied on the definition provided in the clean air act.

- The use of deterministic verses probabilistic values is an underlying philosophical issue that should be dealt with more generally in the report.

- The report overlooks the value of epidemiologic or field data in demonstrating the presence or absence of a cause-effect relationship that can provide a basis for prioritizing or recognizing the efficacy of any management strategy.

- The Agency needs to consider alternative approaches to considering the effects of mixtures on ecosystems which do not rely solely on a molecular mechanism of toxicity to individuals. Most chronic exposures of chemicals to ecosystems are mediated through some ecologically or physiologically significant process that governs fitness (e.g. photosynthesis, respiration, reproduction). Consideration of the effects of mixtures on critical processes of fitness may be worth developing by the Agency.

- The Agency alludes to a more sophisticated level of analysis in higher tiers of the risk assessment, but the document is distressingly short on details. Such detail needs to be added to make the report complete.

- The report overplays the role of conceptual models. These are largely qualitative. There are ecological simulation models that can be used to explore the effects of pollutants on ecosystems. Such models should be recognized and the Agency should explore their use more fully in planning the residual risk assessment program.

Charge 3-4. Health

- The Agency should significantly expand on the issue of data needs for conduct of residual risk assessments in the report and acknowledge the widespread data limitations. As well, the report should contain a discussion of the data collection and research needs and suggest mechanisms by which the data gaps can be filled.

- The data gap issue is so fundamental to the process it should be highlighted in a separate section. A Matrix of Hap's against the data needs should be tabled in this section and methods for prioritize actions be provided.

- The current draft is too limited in its discussion of the Human health problems associated with exposure to HAP's

- The Agency focus on uncertainty in the report is limited to the inputs to modeling efforts. Other possible source should be highlighted. It is important to consider and employ methods for the quantification of both variability and uncertainty

- Dispersion models in general lack validation. The precision (i.e. lack of unexplained variability) and accuracy (i.e. lack of systematic bias) in model predictions should be quantified and considered as a source of uncertainty when performing exposure and risk assessments.

- It is critically important to clearly define risk characterization endpoints prior to analysis and the data selected for the assessment should be at the same temporal and spatial scale as the risk characterization endpoint.

- It is important for the Agency to proceed with simplified screening procedures as a basis for focusing the activities of the Residual Risk Assessment program.

- The Agency should clearly state in the report that there is uncertainty associated with the use of either the MEI or MIR.

- Assessing the unintended consequences of risk management actions is difficult do to the potential for overlapping uncertainties among the predicted risks. Probabilistic methods, if properly employed , can help provide an indication that one risk is higher than another.

Charge 3-4. Ecology

- Although the report identifies the types of data that are used in effects and exposure characterization (Sec. 3.3.2), there is no assessment of the availability of such data for the 188 HAP's. The report should contain a clear statement of the data gaps and how that will provide limitations in assessing residual risks to ecosystems.

- The Agency alludes to a new approach it is developing, but gives no details. Such details should be given even if couched in guarded terms as a developing methodology.

- The report is missing reference to significant sources of data and methods for assessing risks to the environment, such as the use of quantitative models of the ecosystem, the emerging field of ecological economics and the field of ecological epidemiology.

- The report should recognize the paucity of good ecotoxicological benchmarks for environmental effects from exposure to HAP's. In particular key receptor taxa such as plants, birds and wildlife lack relevant data for inhalation and dermal routes of exposure.

- The report should reflect the current state of the art for Ecotoxicological benchmarks and the Agency may want to consider developing its own benchmarking methodology and benchmarks for HAP's to be used in the residual risk program.

- The report should clearly state what data the Agency is planning to use in performing Residual Risk assessments for ecosystems.

Charge 5

- The discussion of how risk managers will utilize the risk assessments in making decisions should be expanded

- The need to validate models should be emphasized in the report. Such models will be relied on heavily for estimating missing data, possible exposures and the resulting effects. Validation will be key to let managers understand the level of uncertainty in the risk estimates.

- The role of stakeholders in the residual risk program needs to be defined in the report.

- The addition of appropriate references would greatly enhance the report.

- The report would benefit from a generic process map that provides a representation of how the Agency intends to prioritize HAP's for analysis, how tiers will function and how data needs will change between screening level and definitive levels of risk assessment.

- The iterative screening technique would be strengthened by greater specification of the procedure.

APPENDIX A-2

Comments of Tom Burke on the Residual Risk report to Congress

Overview and General Considerations

The report provides an overview of "work in progress" on a extraordinarily complex regulatory program which will shape air pollution policy for decades to come. It integrates a broad range of public health, regulatory, technical, and social considerations to provides a framework for proceeding with the Act. The review of the report should consider that the mandates of the CAA go beyond our current abilities to understand and predict health and ecological endpoints. EPA and State regulatory agencies are faced with the difficult challenge of addressing residual risks which are currently not understood. From a public health perspective the most telling statements of the report are found in Section 4.1. Public Health Significance. "Currently the data are not available to conduct an analysis to determine the public health significance for air toxics. In addition, EPA has not completed any residual risk analysis for specific source categories". Clearly, there is a critical need for strengthening the public health basis for the residual risk program.

The document should be considered a framework for moving forward, which is necessarily flexible (perhaps vague) to accommodate an inclusive decision making process. Under the approach stakeholders will have unprecedented involvement, and a rigid prescriptive approach would have little chance of success. It should also be recognized that implementation of the program will happen at the state and local level, therefore flexibility is essential to address and manage risks on a site-specific basis.

The limitations of current data on residual risks, particularly actual population exposures and public health implications are daunting. The Report to Congress presents a pathway for EPA to act based upon available information while identifying data needs for more detailed risk assessments. The report does not provide specific recommendations or approaches for filling these data gaps. Addressing the gaps is essential to successful implementation.

Little consideration is given to developing the technical capabilities of state and local regulators and public health officials. The Report details an iterative process which is beyond the current financial and technical resources of local air quality regulators. Recommendations for building local capacity to evaluate and address residual risks should be included.

Work Assignment

Has the Residual Risk Report to Congress properly interpreted and considered the technical advice from:
b. The 1997 report from the Commission on Risk assessment and Risk Management (CRARM) in developing its risk assessment methodology and residual risk strategy?

The Report is heavily influenced by the recommendations of CRARM, and for the most part does an effective job of integrating its recommendations into the framework. The tiered approach is consistent with the approach recommended by CRARM, providing a practical approach to evaluating risks and addressing those which are most important. *(In SAB report the flowcharts of the tiered approaches from both CRARM and EPA Report should be included side by side to demonstrate similarities and differences.)*

To assure that the CRARM recommendations were considered, Section 5.3.5. of the Report lists each and describes how they were addressed. For the most part EPA is "in the process" of developing strategies to address each point. Potential differences in implementation and interpretation are listed below.

Characterize and articulate the scope of the national, regional, and local air toxics problems and their public health and environmental contexts.

EPA is obviously just beginning this process, particularly regarding the public health and environmental contexts. The CRARM calls for a broad public health approach which examines the actual health impacts on the effected communities, and considers the residual risk in the context of the health status of the population. While the Report mentions the collection of population health data and potential integration of epidemiological approaches, no specific methods are detailed, and no commitment to tracking the health status of the population is made. The proposed program is largely driven by animal based cancer bioassays to estimate public health context. Without developing a more detailed approach it is not possible to determine if EPA is implementing this CRARM recommendation. More importantly, without a specific strategy to evaluate population health status it will be difficult, if not impossible, to determine if the residual risk management makes any difference in the public's health.

At facilities that have upper bound cancer risks greater than one in 100,000 persons exposed or that have concentrations greater than reference standards, examine and choose risk reduction options in light of total facility risks and public health context.

According to the CRARM Report, this recommendation is to develop a flexible bright line that considers local public health impacts and the total facility risk. EPA does not adopt the one in 100,000 approach, opting for the flexible approach of the benzene NESHAP. Specific approaches for non-cancer effects are under development and not specifically detailed in the Report. It is not apparent just how "public health context" will be interpreted. This should be addressed in greater detail in the Report in order to better represent the recommendation of CRARM.

Consider reduction of residual risks from source categories of lessor priority.

EPA interprets this as a mandate to do the "worst first", and considers the Report to address this recommendation. Further consideration of lessor sources should be included to address the management of high background risks, to protect populations with high aggregate or cumulative risk, or to consider the public health of sensitive populations.

Other issues to consider

Stakeholders - the cornerstone of CRARM is stakeholder involvement. The Report needs to be more specific in identifying who the stakeholders are and how they will be engaged throughout the process. This should include those at the national, state, and local levels.

Epidemiology - the Report is generally negative regarding the application of epidemiology to the evaluation and management off residual risks. As mandated by the law, EPA should consult with the public health community to develop a public health based surveillance system to track population health and provide a continual public health context for residual risk management. If EPA concedes from the start that the public health benefits of the program are not measurable is the cost worth it?

Linkages between ecological health and human health. These are not addressed in the report. Human health is a powerful environmental indicator. Common aspects of ecological risk assessment and public health surveillance should be described.

Evaluation - How will we know the approach is working? Key indicators of success should be identified and methods for tracking them included in the Report.

Background Risk - In order to provide an "ample margin", background risk should be considered. More detail is necessary to understand the EPA approach for both health and ecological endpoints.

Sensitive subpopulations - The Report does not specifically address how such populations will be considered in the risk assessment process.

APPENDIX A-3

Comments on Draft Residual Risk Report to Congress
Prepared for Residual Risk Subcommittee of the US EPA Science Advisory Board's Executive Committee

Prepared by:

H. Christopher Frey, Ph.D.
Assistant Professor
Department of Civil Engineering
North Carolina State University
Raleigh, NC 27695-7908
(919) 515-1155
(919) 515-7908 (fax)
frey@eos.ncsu.edu

August 5, 1998

Introduction

This document contains my post-meeting comments (pages 1-3), my pre-meeting comments (pages 4-11), Appendix A with a brief literature review on probabilistic methods (pages 12-14), Appendix B with my comments previously submitted to EPA regarding the ASPEN modeling approach mentioned in the draft Report (pages 15-21), and Appendix C with a draft summary of a recent EPA workshop, which I chaired, regarding uncertainty analysis (pages 22-30). I have added a few minor clarifications to my pre-meeting comments. Thus, I recommend that this file be used as a basis for preparing the committee report, and that the file submitted prior to the meeting be discarded.

In my post-meeting comments, I endorse and expand upon some of the general points that were made at the Residual Risk Subcommittee meeting on August 3. These comments are in addition to my pre-meeting comments.

Post-Meeting Comments

The Report to Congress should convey the limitations, data collection needs, and research needs associated with risk assessment. The Report should give clearer context for the current state of practice of risk assessment, and provide a roadmap for what additional information, data, models, etc. would be needed to comply in full with the current requirements of the Clean Air Act regarding residual risk assessments. Where requirements of the Act appear to be optimistic or unrealistic, it would be useful to give an indication of data collection and

research activities that would be needed in order to proceed with the conduct of the assessments.

There is a strong need for more data and for methods to prioritize data collection in support of the residual risk assessment activities. It should be clearly noted in the report that in many cases, data of sufficient quality and quantity are not available at this time to fully support the risk assessment effort. Methods for prioritizing data collection include the use of models. For example, sensitivity analysis and probabilistic analysis can be used to help identify key sources of uncertainty, associated with lack of data, upon which risk estimates may be highly dependent. These key sources of uncertainty can be targeted for additional data collection as needed. This approach, in combination with valuation of the cost of collecting the data, provides a systematic method for setting ongoing research agendas. As new data are collected, the need for additional data can be re-evaluated and resources can be retargeted as needed to the next most important key sources of uncertainty.

A specific area where data are severely lacking is regarding emission rates, emission inventories, and ambient air quality data for the 188 HAPs. Emission measurements for HAPs can be expensive and difficult, which accounts in part for the lack of data. Because HAPs have only recently (compared, for example, to criteria pollutants) become the subject of regulatory scrutiny, databases are only now being developed, and typically are incomplete. For example, HAP emissions are, in general, poorly characterized. Data tend to be available only for a small subset of the 188 pollutants, and the quality of data varies greatly among the 170 source categories. To support both chronic and acute health risk assessments, it is necessary to measure HAP emissions for long time periods using short sampling times (e.g., years worth of hourly or daily data). In the absence of such data, many assumptions (judgments) will be needed in order to make estimates of emissions for averaging times that are appropriate to health and ecological risk assessments. The use of judgment is inherent in any risk assessment process, and should be recognized and made as transparent as possible to facilitate peer review.

In order to develop emissions estimates for use in risk assessment, it will be necessary to consider not just emission factors or emission measurements at a representative set of facilities for each of the source categories, but also to consider the activity levels of the emission sources within the geographic scope of each assessment in order to develop an emission inventory. An inventory is typically conceptualized as the product of an emission factor (e.g., mass emission of a pollutant per unit of activity associated with the release of the pollutant) and an activity factor (e.g., the number of units of activity), summed over all emission sources. Data on activity factors can be difficult to obtain. For many source categories, activity is highly variable, especially over short averaging times. In addition, because activity data may be difficult to obtain, there is often substantial uncertainty regarding activity levels. Thus, the collection of activity data may become an important priority for improvement of risk assessments. The importance of the collection of emissions and activity data can be assessed via sensitivity and probabilistic analyses as noted previously.

The development of risk estimates will likely rely heavily on the use of dispersion models. It should be noted that the typically employed Gaussian-based dispersion models are considered to be precise to no better than plus or minus 30 percent and are only appropriate for evaluation of short-range transport (less than 50 kilometers). The preliminary results from the ASPEN modeling effort suggest that uncertainties may be far greater than plus or minus 30 percent. There is a lack of validation of such dispersion models in many cases. Furthermore, it is not likely that the dispersion of all HAPs can easily or appropriately modeled using Gaussian plume models, due to their chemical reactivity and/or physical characteristics. The comparison of air quality modeling predictions, which will included uncertainties associated with emissions estimates, stack parameters, meteorological scenarios, and the structure of the models themselves, with measured ambient monitoring data, is an important means to provide insight into the precision and accuracy of the models. Thus, efforts should be continued and expanded regarding the collection of ambient HAPs measurements. Precision refers to lack of unexplained variability in model predictions, whereas accuracy refers to lack of systematic bias in model predictions. The precision and accuracy of dispersion models should be quantified and considered as a source of uncertainty when performing exposure and risk assessments. Results from the ASPEN effort may be useful in this regard, although they suggest as previously noted that the precision of the dispersion models are rather poor.

Along the lines of continued and additional measurement of ambient HAPs concentrations, it is important to develop a sound basis for estimation of background levels. At this time, the estimation of background levels is highly uncertain and perhaps even speculative in many cases. A program of additional measurements should be considered as a means to improve the database and reduce uncertainty regarding estimation of background levels.

It is critically important to clearly define the risk characterization endpoints prior to performing a significant number of analyses. In fact, the definition of endpoints is needed early on in order to help anticipate data collection and research needs in support of the residual risk program. For example, the evaluation of various health and ecological endpoints will have implications for the temporal and spatial characteristics of each assessment. For acute endpoints, data based upon short averaging times (e.g., hourly, daily) will be required for all assessment inputs. For chronic endpoints, data based upon long averaging times will be required. As noted elsewhere, for example, emissions data may in some cases be available for a convenience sample of short averaging times (e.g., daily), but collected only over a short testing program (e.g., only for a few days). In this example, temporal patterns in emissions (e.g., seasonal variations, autocorrelations) would not be likely to be revealed. Thus, emissions data collected over a short duration for only a limited number of short time periods may not be adequate for supporting acute risk assessments, nor would it be a sound basis for making chronic risk assessments. The geographic scope of assessments also has important implications for data collection. If localized, acute health effects are to be studied, then highly location-specific data may be required. In contrast, if chronic effects that may result from longer range transport are of importance, then "representative" regional or national average

data may be sufficient. Variability and uncertainty tends to increase as the averaging time or geographic scope of a study decreases.

Because it is unlikely that all data gaps will be filled prior to the development of residual risk estimates, it will be especially critical to consider and employ methods for the quantification of both variability and uncertainty. These methods are more fully addressed in my premeeting comments.

In order to more realistically manage the residual risk requirements, it will be necessary for EPA to prioritize the focus of the assessment effort. Prioritization may be easily accomplished by screening the list of 188 HAPs to identify those that are least active in terms of human and ecological health effects, and to focus initially upon those that appear to pose the greatest threats. Similarly, EPA should prioritize the 170 source categories not merely based upon the timing of implementation of MACT standards for those categories, but based upon screening-level assessments of which source categories may pose greater residual risks than others. It is important that EPA proceed early on with simplified screening procedures as a basis for focusing the activities of the residual risk assessment program. As new data become available, the screening studies should occasionally be revisited to make sure that no important HAPs and/or source categories are overlooked.

The report should be careful to convey that uncertainties tend to be greatest at the extreme tails of distributions, such as for distributions of the variability of exposure or risk over a population of exposed individuals. Therefore, measures such as MIR are likely to be highly uncertain compared to average population risk characteristics. The uncertainties in risk estimates typically span orders-of-magnitude, when all sources of uncertainty are accounted for (including uncertainty in the dose-response relationship).

Because uncertainties in risk assessments are typically large, there is a special challenge for the evaluation of unintended consequences. When comparing two risks, it can be difficult or impossible to determine which one is really higher, because both may be uncertain by orders-of-magnitude and have overlapping uncertainty ranges. Probabilistic methods, if properly employed, can help provide an indication of the likelihood that one risk is really higher (or lower) than another risk. However, it should be expected that the results of such assessments may not be definitively conclusive in many cases.

Pre-Meeting Comments

Submitted: August 2, 1998

My comments focus mostly upon the uncertainty aspects of the Report.

In reference to the discussions of control technologies and pollution prevention measures: In general, it is important to consider variability and uncertainty in control technology efficacy and cost, in addition to the other sources of variability and uncertainty in exposure and risk

assessments. The probabilistic methods described for exposure and risk assessment are typically general enough for application to technology assessment problems (e.g., see Frey and Rhodes (1996), Frey *et al*. (1994), and Frey and Rubin (1998) for examples of probabilistic technology assessments).

I served as a reviewer for the ASPEN modeling approach described on p. 35 and will provide a copy of my comments on that as an attachment.

The most recent presentation that I heard regarding TRIM, at the Society for Risk Analysis annual meeting in December 1997, was indicative of an incomplete approach for quantification of variability and uncertainty, in contrast to the assertions on pages 36 and 41 of the Report. Essentially, it appeared as if both variability and uncertainty were to be combined in one dimension of probabilistic analysis. This situation may have changed; however, I would be cautious about the use of TRIM until it has undergone external peer review. The Report should state that the use of any of the approaches described here, such as ASPEN or TRIM, will be considered only after these approaches have undergone sufficient peer review.

[Addendum: based upon discussions with OAQPS personnel in attendance at the SAB meeting, my understanding is that an improved approach for distinguishing between variability and uncertainty is being considered for TRIM. However, this proposed capability for TRIM should receive peer-review, as I understand is intended.]

The discussion in Section 3.1.4 regarding Risk Characterization, and specifically regarding uncertainty and variability, is quite reasonable.

p. 55. It should be noted that direct measurement of HAP emissions is not a panacea, in the sense that one should not expect highly accurate and precise emissions estimates even if some measurement data are available. HAP emissions can be highly variable over time and from source-to-source, even within a source category. In addition, measurement of HAPs can be fraught with many difficulties, especially regarding sampling of the stack gases. The precision of measurement methods is probably typically no better than plus or minus 25 percent, but there are also uncertainties regarding the accuracy of some methods applied to some compounds.

p. 56 (1st full paragraph). Some care needs to be taken with terminology. The term "short-term" as applied to emissions typically has the connotation of a short term stack test (e.g., a three day test). Such data could not reliably be used to make estimates of emissions "over a range of release times," as suggested in the Report. More likely, the paragraph was intended to convey that if emissions data were collected over a long time period using a short sampling time (e.g., a year's worth of hourly emission data), then it would be possible to make emission estimates for averaging times from one hour to one year (for example) for that particular source. Even this would be true only if there was no inter-annual variability and as long as

any seasonal variations were appropriately characterized. Issues of temporal autocorrelation in emissions would also have to be evaluated. Since HAP emissions are not typically measured using continuous monitoring, such data are not likely to be available.

p. 67. It should be anticipated and stated that there is uncertainty regarding both the MEI and the MIR. It is appropriate to constrain the MIR to be representative of an actual person, rather than a fictitious "porch potato" or resident in the middle of a lake. However, it is also important to consider the 1992 Exposure Assessment Guidelines and include the notion of a high end exposure and a mid-range exposure in assessments beyond the screening stage. These can easily be inferred from the results of probabilistic analyses. The range of uncertainty for the MIR is likely to be very large compared even to the risks associated with high-end exposures (e.g., around the 90th percentile). It is not realistic to expect any method to be able to make a precise prediction of the MIR, and this should be clearly stated in the Report.

The approach to be taken for Margin of Exposure analyses should be subject to external peer review at such time as the approach is available in draft form.

Section 4.2.3

In response to my charge to be the lead on uncertainty, especially section 4.2.3, I offer the following comments.

First Paragraph of Section 4.2.3

The first paragraph requires some reorganization and better structure. There are broader sources of uncertainty than are mentioned here. The following should be mentioned:

a) Uncertainty in selection of representative scenarios, including pollutant sources, transport, exposure pathways, exposed populations, etc.

b) Uncertainty in the structure of models used to represent a given scenario

c) Uncertainty and variability in the inputs of the model(s).

The report tends to focus only on this latter source of uncertainty. However, the first two may be more important in many cases. The first one can be addressed by analysis of multiple scenarios. The second one can be addressed by analysis using more than one modeling approach. The third can be addressed using probabilistic methods as described in the report. Some would argue that the first two can also be addressed by probabilistic methods.

Second Paragraph of Section 4.2.3

In the second paragraph, there seems to be a distracting discussion of the definition of "uncertainty analysis", which is posed as a term that has little meaning and that is misleading. While the points made in the second and third sentences have some validity, they are not particularly important. Furthermore, they can be easily addressed by using terms such as "sensitivity and probabilistic analysis", which encompass many types of analyses and also encompass analysis of both variability and uncertainty.

Third Paragraph of Section 4.2.3

The distinction between variability and uncertainty has roots prior to the EPA (1997a) report that is cited here. To add credibility to the distinction, earlier reports and papers should be cited, including peer-reviewed publications. I have prepared a brief appendix to these comments providing a literature review (from my recent peer-reviewed papers) on this subject, which I offer for consideration and inclusion in the revised Report.

The key questions listed at the bottom of page 90 are generally good. The first question leaves open the possibility of uncertainty in models, which is often an important issue. To this should be added uncertainty in scenarios that have been selected for analysis.

Page 91

It is encouraging to see the issues of uncertainty and variability addressed from both a risk assessment and a risk management viewpoint, without any negative assumptions regarding the putative inability of risk managers to deal with uncertainty, as indicated in the CRARM report.

To the list of "major documents" on page 91, I would add the following:

Summary Report for the Workshop on Monte Carlo Analysis, EPA/630/R-96/010, September 1996.

This report provided a technical basis for the 1997 documents (Policy for Use of Probabilistic Analysis in Risk Assessment, and Guiding Principles for Monte Carlo Analysis) and is the product of an EPA-sponsored workshop in which many experts outside of the Agency were participants. The summary report provides additional details regarding alternative methods and case studies that will be useful to many people.

It also should be noted that the Risk Assessment Forum convened a workshop in New York City in April 1998 on "Selecting Input Distributions for Probabilistic Analysis." The workshop was comprised of experts, mostly from outside of EPA. The summary report from this workshop has undergone review and should be available soon. The EPA contacts are Steve Knott and Bill Wood. If possible this summary report should be cited. For your convenience I will attach my summary of the workshop (I was the chair), which is in draft form.

Pages 92-93

The discussion on pages 92-93 regarding several approaches for addressing variability and uncertainty provides useful information. However, more context is needed prior to the discussion of each alternative. Specifically, the notion of a tiered approach to sensitivity and probabilistic analysis should be introduced, as discussed on p. 5 of the EPA (1997) Guiding Principles for Monte Carlo Analysis. The notion of a tiered approach is described in more detail in the EPA (1996) Summary Report, on pp. 3-3 to 3-4, and pp. E-3 to E-8.

In the discussion of a tiered approach from the 1996 Summary Report, it is noted on p. E-5 that there are "five factors that determine the precision or reliability of a health impact assessment [these factors may also be applicable to ecological impact assessments]: (1) specification of the problem (scenario development); (2) formulation of the conceptual model (the influence diagram); (3) formulation of the computational model; (4) estimation of parameter values; and (5) calculation and documentation of the results including uncertainties." The proposed tiered approach to analysis of variability and uncertainty involves four tiers:

1. Single-value estimates of high-end and mid-range risk
2. Qualitative evaluation of model and scenario sensitivity
3. Quantitative sensitivity analysis of high-end or mid-range point estimates
4. Fully quantitative characterization of uncertainty and uncertainty importance

While these are not the only possible tiers, they are suggestive of an approach which may begin with evaluation of a small number of alternative scenarios, coupled with qualitative discussions of uncertainty, and then may proceed through more elaborate sensitivity analyses, perhaps culminating in a "two-dimensional" simulation of both variability and uncertainty for alternative scenarios and model formulations.

There seems to be some confusion over variability and uncertainty as indicated by the text on pages 92 and 93. Most of this text appears to be focused upon uncertainty analysis, but implies that a great deal of data are required in order to do any of the suggested types of quantitative analyses. This is illogical. Uncertainties are typically greatest when data are limited or irrelevant to the problem at hand. Thus, it may be difficult to characterize variability in such situations and it is especially important to attempt to characterize uncertainty.

The discussion of the "Multi-Scenario Approaches and Limited Sensitivity Analysis" on p. 92 contains a factual error. The statement that sensitivity and uncertainty analyses are "often limited to only those variables for which data are available (which is true of all quantitative treatments of uncertainty)" is wrong. *Uncertainty* is typically greatest when data are not available, and methods for dealing with uncertainty in such situations have been developed and applied. Such methods are discussed in the EPA (1996) Summary Report, as well as in the peer-reviewed literature, books, reports, etc. For example, there are several protocols

which have been developed for eliciting expert judgments regarding uncertainty in the absence of directly relevant data. One of the most widely reported protocols is one developed in the 1960s and 1970s at Stanford and the Stanford Research Institute (Spetzler and von Holstein, 1975, and as discussed by Morgan and Henrion 1990, Morgan *et al.*, 1980, and Merkhofer, 1987). The Stanford/SRI protocol involves five steps. Similar protocols have been developed by others. In addition, there are methods for combining judgment and data based upon "Bayesian" approaches, as briefly described in EPA (1996) and elaborated upon elsewhere.

The mistaken notion that uncertainty analysis is data intensive raises many issues, which have been addressed at the two EPA-sponsored workshops previously mentioned and elsewhere. Briefly, directly relevant data are rarely available. Therefore, considerable judgment goes into the selection of data as the basis for specifying input assumptions in a model. The selected data are typically merely surrogates of some quantity (e.g., activity data for a population similar to, but not the same as, the one under study). Thus, there is a subjective element already embedded into the selection of input assumptions, whether for a point estimate or a probabilistic assessment. The April 1998 workshop delved into issues of representativeness of data and distributions in some detail. The panel generally considered that the objective in specifying values or distributions for inputs to a model was to achieve "adequacy" with respect to the purpose of the particular analysis. The notion of "adequacy" pertains to the population, temporal, and spatial characteristics of the study, as well as the "who, what, why, when, where, and how" of the endpoint of the assessment. In many cases, it is necessary to use surrogate data. Furthermore, it is often necessary to use "plausible extrapolation" methods when data are limited, especially for the purpose of characterizing higher percentiles for a given model input.

When directly relevant, randomly sampled data are not available, then judgment is inherent in the process of specifying inputs to a model. This is precisely the type of situation in which there are typically significant amounts of uncertainty. Expert judgment must be an acceptable basis for estimating uncertainty; otherwise, it is certain that uncertainty will be underestimated.

The same paragraph also mentions "combinations of variable values that are used to derive the various risk estimates may not be physically plausible." This issue received some attention at the April 1998 workshop. It is possible to avoid this by proper specification of the range of values for each model input and proper specification of any correlation structures among the inputs. However, it is also the case that model outputs are typically most sensitive to only a few of the model inputs. Thus, if there are implausible combinations of values to which the model output is not sensitive, then it is not likely that the model results would be affected. Furthermore, it is also not necessarily the case that an extreme value for one model input is associated with an extreme value of a model output. For example, in a probabilistic simulation, the upper tail of the distribution of a model output may be due to various combinations of values of the model inputs, not necessarily a worst case combination of all of the input values. Therefore, the concern over implausible combinations of model inputs

is a relatively minor point, especially at the level of "limited sensitivity analysis", where is it is usually relatively easy to choose a small number of plausible combinations of model inputs.

The paragraph on "Systematic Sensitivity Analysis" has a curious and inappropriate start with "When sufficient data are available...". Again, the whole point of uncertainty analysis is to characterize the implications of lack of knowledge. Lack of knowledge is often greatest when data are limited or non-representative. In such situations, one might argue that data are not "sufficient". However, if a policy decision must be made regardless, then it is still useful to develop sensitivity ranges based upon analogies with surrogate data sets.

The techniques mentioned in this paragraph are usually appropriate only after one has developed a good model and run it for many case studies. For example, correlation analysis presumes that there are sets of model inputs and outputs that can be analyzed statistically. In practice such model input and output data sets most likely would be developed using probabilistic analysis techniques, such as Monte Carlo simulation. Thus, the ideas here are really more appropriate for evaluation of the importance of inputs to a probabilistic analysis, and in practice would not typically be a separate tier of an uncertainty analysis prior to probabilistic analysis. A counter example to this would be the use of regression analysis or response surface model as part of the development of an integrated assessment model. In such cases, a simplified model is developed based upon a more complex model based upon systematic sensitivity analyses of the complex model. The simplified model can then be coupled to other simplified models that represent other portions of a scenario (e.g., alternative transport and fate pathways). The entire integrated assessment model can then be used for limited sensitivity analysis or perhaps for probabilistic analyses. This approach was employed, for example, in an integrated assessment of acid deposition, resulting in a model called the "Tracking and Analysis Framework" (TAF). TAF contains reduced form versions of more detailed models, such as for regional transport and deposition of "acid rain" species. The simplified models for emissions, transport, effects, and valuation were combined in an integrated probabilistic assessment model. (Project details are available at http://209.24.95.115/taflist/)

The techniques for systematic sensitivity analysis are not necessarily "very difficult to interpret", nor are they necessarily more resource-intensive than, for example, probabilistic methods. Response surfaces, for example, can often be very informative. The variation of a model output (e.g., exposure, risk) as a function of two inputs can easily be displayed using a three dimensional graph. The sensitivity of model outputs to many model inputs can be conveniently summarized using sample or rank correlation coefficients, partial rank correlation coefficients, or standardized regression coefficients, or with other measures. However, as previously noted, often these latter types of sensitivity measures are calculated based upon the results of a probabilistic analysis.

Techniques missing from the discussion, which can be very useful, are interval analysis and probability bounds methods. These methods allow for relatively simple characterization of

ranges of values for each model input, and also allow for consideration of all possible correlation structures between the inputs. However, because these methods typically do not make use of all of the information known regarding model inputs, they can produce very wide ranges for model outputs. While these techniques are conservative in overpredicting the model output ranges, they may not be particularly informative. Bounding methods are mentioned in the EPA (1996) summary report.

The paragraph on "Monte Carlo Simulation and Related Probabilistic Methods" again fixates on the notion of data intensity as a prerequisite to probabilistic analysis. This is an unrealistic requirement and will serve to stifle any analyses beyond a simple and misleading point estimate. While it is certainly desirable to have a large amount of randomly sampled directly relevant data, it is rarely the case that such data are available. Therefore, there is often a limited database from which to characterize variability in a model input. Fortunately, there are methods for simultaneously characterizing both variability and uncertainty for small data sets (e.g., see Frey and Rhodes, 1996; Frey and Rhodes, 1998; Burmaster and Thompson, 1998; Frey and Burmaster, 199x, etc.). Furthermore, in the context of a particular assessment it is often possible to identify and model more than one source of uncertainty (e.g., random sampling error, lack of precision and accuracy of a measurement method, etc.). It is often the case that variability may have to be extrapolated beyond the range of available data. Here again, methods such as bootstrap simulation, likelihood estimation, and others can be used to quantify the range of uncertainty in the tails of a distribution that has been extrapolated beyond the range of observed data. Therefore, such methods are not *data* intensive in the sense of requiring large data sets for each model input; instead, they may be *computationally* intensive in terms of the number of alternative values that are simulated for each model input as part of a probabilistic simulation.

It is not appropriate to make a blanket generalization that "results depend strongly on the availability of information or the resources to gather information." This would only be true for the most sensitive inputs. It would not be true for insensitive inputs.

The sentence "the outputs of simulation models may be difficult to interpret for stakeholders and risk managers accustomed to discrete risk estimates" seems a bit unfair. This will depend on who the stakeholders and risk managers are and on how the model results are presented. Issues of variability should be relatively straightforward to communicate. Quite simply, not everyone has the same exposure or risk. It is possible to present a few alternative realizations from the probabilistic analysis to illustrate this. For example, Individual A has a low exposure because of a particular activity pattern compared to Individual B. Issues of uncertainty should also be possible to communicate. For example, for any one individual we do not know exactly what their exposure or risk is, because it is impractical to measure each person's activity patterns and we do not have complete knowledge of the means by which exposure to a particular chemical for a particular time period at a particular concentration results in a given health effect. Thus, there is uncertainty regarding each individual person's exposure and risk. Because there are uncertainties for all individuals, we are also uncertain as to what the

risks are to the "average" member of the population, to "highly exposed individuals", etc. Furthermore, we are uncertain regarding the number of cases of a particular health affect among the exposed population. Specific examples can be given for each of these as needed.

It is quite true that "simulation modeling can rarely be used to capture all courses of variability and uncertainty quantitatively." Here it is worth adding that issues of structural uncertainty associated with scenarios and models can be addressed through evaluation of alternative cases.

Not mentioned in the discussion of uncertainty is the issue of model validation. Many of the models that are used in exposure and risk assessment are poorly validated, if at all. In principle, the precision and accuracy of a model should be known and incorporated into the probabilistic analysis. It is also typically not necessary to perform thousands of Monte Carlo simulations with a model that may only be precise to plus or minus 50 percent.

"Strategy for Considering Uncertainty in Residual Risk Analyses". This paragraph is generally good, but it would be better to state more clearly what the approach to uncertainty evaluation will be. Rather than say that a tiered approach "will likely be adapted", why not say that a "tiered approach will be adapted". It is okay if the details of the tiered approach are not specified at this time, but it should be clear that a tiered approach is anticipated and expected.

Page 94

Top of page 94. It is valuable to identify key sources of uncertainty, especially when taking a longer term view of the risk management process. Risk management will improve as uncertainties are reduced. Key sources of uncertainty can be identified, based upon probabilistic analysis, and then targeted for additional research and data collection. While it is possible that a "simple multi-scenario approach" may be sufficient in some cases, one should keep in mind that probabilistic analysis is also a "multi-scenario" approach. Once a computational model is formulated and once ranges have been identified for model inputs, it is usually not significantly more difficult to run a probabilistic analysis than it is to do multiple sensitivity analyses. In fact, it may be easier, depending upon the software.

"Uncertainty and the Management of Risks"

This paragraph has a rather strange introduction. The second sentence seems to have a message between the lines which appears to this reader to be overly negative. It would be fair to say that analysts have developed new methods for a fuller quantitative characterization of variability and uncertainty. These methods pose new challenges for the development of summaries of results for use by decision makers. In part, these challenges are because of the richness of the information provided by the new methods. I would then delete the first seven lines of this paragraph.

The notion of the availability of specific control options strongly suggests that there be analyses of alternative scenarios regarding implementation of controls, and that these be done probabilistically so as to allow for evaluation of uncertainty in the efficacy and cost of the technologies. These uncertainties can span orders-of-magnitude. For example, EPA and DOE studies regarding mercury control costs for electric power plants differ by an order-of-magnitude, based upon a recent presentation at the U.S. DOE's Federal Energy Technology Center (Brown *et al*., 1998).

Last paragraph of Section 4.2.3

There has to be a careful distinction between the notion of "complexity" as faced by analysts in performing risk assessments and probabilistic analyses, versus the "complexity" faced by the decision maker in interpreting the results of such analyses. While it is true that analysts may have to grapple with many difficult problems and decisions, it is possible to summarize the most important findings and caveats in a compact form for consumption by decision makers. The nitty gritty details of an analysis can always be given in an accompanying report (and should in any event be subject to scientific peer review prior to use in decision making). We should not expect decision makers to conduct a detailed technical review of an assessment; that should be done via peer review, preferably with scientists external to EPA. However, at least some decision makers have in the past expressed a preference for probabilistic presentations of risk information. Bloom et al. (1993) conducted a focus group study of several EPA decision makers and evaluated their preferences for various methods for communication of uncertainty information. Perhaps surprisingly, many expressed a preference for one of the more detailed forms of communication Ñ a cumulative distribution function.

Information regarding variability and uncertainty can often be presented to stakeholders in more of a narrative format as suggested previously. Although the tone of this paragraph is overly negative, it does nonetheless appear to take a constructive approach to dealing with the issues of presentation and communication of uncertainty information. Specifically, it is encouraging that the Report indicates that efforts will be continued to improve transfer of information.

Brief Literature Review on "Variability and Uncertainty".

Based upon Frey and Rhodes (1998) and Frey and Burmaster (199x).

While there has been considerable work in the quantification of uncertainty in human health risk assessments, in the last five years or so there has been increasing attention to the distinction between "variability" and "uncertainty." A diversity of definitions regarding variability and uncertainty can be found in: Bogen and Spear, 1987; Frey, 1992; Hoffman and Hammonds, 1994; MacIntosh *et al*., 1994; McKone, 1994; Frey and Rhodes, 1996; Hattis and Barlow, 1996; Price *et al*., 1996; and others. Variability refers to diversity among members of a population. For example, there are differences in exposures to chemicals among different members of a population of people. Uncertainty refers to lack of complete knowledge

regarding the true value of a quantity. For example, there is usually lack of knowledge regarding the true values of exposures for any given member of a population and, therefore, regarding the distribution for variability among all members of an exposed population over space and time. Both variability and uncertainty may be described using probability distributions.

A number of approaches to characterizing variability and uncertainty in the inputs to exposure and risk models have been developed. In general, characterizations of both variability and uncertainty in a model output (*e.g.*, exposure, risk) must rely upon specification of both variability and uncertainty in the model inputs and upon a method for propagating these inputs through the model. Bogen and Spear (1987) present a mathematical framework for estimating variability and uncertainty in model outputs. Frey (1992), Hoffman and Hammonds (1994), MacIntosh *et al*., (1994), Frey and Rhodes (1996), Cohen, Lampson, and Bowers, (1996), Frey and Rhodes (1998), and others have employed numerical methods to propagate both variability and uncertainty through a model. These methods have typically employed Monte Carlo or related sampling techniques (*e.g.*, Latin Hypercube sampling) in two separate dimensions. One dimension is devoted to uncertainty, while the other is devoted to variability. Bogen (1995) presents an approximate method for propagating both variability and uncertainty through models based upon discretization of input distributions. Rai, Krewski, and Bartlett (1996) present an approximation method based upon the use of Taylor series expansions. A numerical simulation method is described by Frey and Rhodes (1996) for propagating both variability and uncertainty through a model.

Burmaster and Thompson (1998) have employed a likelihood-based method for estimating sampling distributions. Frey and Burmaster (1998) compare bootstrap and likelihood-based approaches to characterizing both variability and uncertainty with respect to three data sets and three types of frequency distributions (*i.e*. Normal, Lognormal, and Beta).

The development of input assumptions for second-order random variables may be based upon expert judgment and/or the analysis of data. For example, expert judgment has been employed in a variety of analyses (e.g, Hoffman and Hammonds, 1994; NCRP, 1996; Barry, 1996; Cohen *et al.*, 1996). Statistical techniques based upon the analysis of data which have been applied to second-order random variables include the bootstrap method (e.g., Frey and Rhodes, 1996) and maximum likelihood (MLE) methods (Burmaster and Thompson, 1998). After the inputs to a model have been specified as second order random variables, a variety of methods may be used to propagate both variability and uncertainty through the model to estimate both variability and uncertainty in the output. These methods include mathematical approaches (e.g., Bogen and Spear, 1987), "two-dimensional" Monte Carlo-based simulations (e.g., Frey, 1992; Hoffman and Hammonds, 1994; and others), and approximation methods based upon discretization of input distributions (e.g., Bogen, 1995) or the propagation of moments using Taylor series expansions (Rai *et al.*, 1996).

References Cited

Barry, T.M. (1996), "Distributions on a Budget," presented at US EPA's Workshop on Monte Carlo Analysis, 14 May 1996, New York, NY .

Bogen, K.T., and Spear, R.C. 1987. "Integrating Uncertainty and Interindividual Variability in Environmental Risk Assessment," *Risk Analysis*, **7**, 4, 427-436.

Bogen, K.T. 1995. "Methods to Approximate Joint Uncertainty and Variability in Risk," *Risk Analysis*, **15**, 3, 411-419.

Brattin, W.J., Barry, T. M., and Chiu, N. 1996. "Monte Carlo Modeling with Uncertain Probability Density Functions," *Human and Ecological Risk Assessment*, **2**, 4, 820-840.

Brown, T., W. O'Dowd, R. Reuther, and D. Smith (1998), "Control of Mercury Emissions from Coal-Fired Power Plants: A Preliminary Assessment of Cost," Presented at Advanced Coal Based Power and Environmental Systems Conference, U.S. Department of Energy, Morgantown, West Virginia, July 21-23, 1998 [proceedings will be available on CD ROM in a few months].

Bloom, Diane L., Dianne M. Byrne, and Julie M. Andresen (1993), "Communication Risk to Senior EPA Policy Makers: A Focus Group Study." Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Burmaster, D.E. and Thompson, K.M. 1998. "Fitting Second-Order Parametric Distributions to Data Using Maximum Likelihood Estimation," *Human and Ecological Risk Assessment*, in press.

Cohen, J.T., Lampson, M.A., Bowers, T.S. 1996. "The Use of Two-Stage Monte Carlo Simulation Techniques to Characterize Variability and Uncertainty in Risk Analysis", *Human and Ecological Risk Assessment*, **2**, 4, 939-971.

Frey, H.C. (1992), Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making, Prepared for American Association for the Advancement of Science and U.S. Environmental Protection Agency, AAAS/EPA Environmental Science and Engineering Fellowship Program, Washington, DC, September 1992.

Frey, H.C., E.S. Rubin, and U.M. Diwekar, "Modeling Uncertainties in Advanced Technologies: Application to a Coal Gasification System with Hot Gas Cleanup," *Energy* 19(4):449-463. (1994).

Frey, H.C., and D.S. Rhodes (1996), "Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example," *Human and Ecological Risk Assessment: an International Journal*, 2(4):762-797.

Frey, H.C., and E.S. Rubin, "Uncertainty Evaluation in Capital Cost Projection," in *Encyclopedia of Chemical Processing and Design*, Vol. 59, J.J. McKetta, ed., Marcel Dekker: New York, 1997, pp. 480-494.

Frey, H.C., and D.S. Rhodes (1998), "Characterization and Simulation of Uncertain Frequency Distributions: Effects of Distribution Choice, Variability, Uncertainty, and Parameter Dependence," *Human and Ecological Risk Assessment*, 4(2):423-468.

Frey, H.C. and Burmaster, D.E. (199x). "Methods for Characterizing Variability and Uncertainty: Comparison of Bootstrap Simulation and Likelihood-Based Approaches," *Risk Analysis*, accepted.

Hattis, D. and K. Barlow (1996), "Human Interindividual Variability in Cancer Risks -- Technical and Management Challenges," *Human and Ecological Risk Assessment*, Volume 2, Number 1, pp 194 - 220.

Hoffman, F.O. and Hammonds, J.S. 1994. "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability," *Risk Analysis*, **14**, 5, 707-712.

Kaplan, S. (1992). "'Expert Information' versus 'Expert Opinions': Another Approach to the Problem of Eliciting/Combining/Using Expert Knowledge in PRA". Reliability Engineering and System Safety. 35:61-72.

MacIntosh, D.L., Suter II, G.W., Hoffman, F.O. 1994. "Uses of Probabilistic Exposure Models in Ecological Risk Assessments of Contaminated Sites," *Risk Analysis*, **14**, 4, 405-419.

McKone, T.E., 1994, Uncertainty and Variability in Human Exposures to Soil Contaminants through Home-Grown Food: A Monte Carlo assessment Risk Analysis, Volume 14, Number 4, pp 449 - 463.

Merkhofer, M.W. (1987). "Quantifying Judgmental Uncertainty: Methodology, Experiences, and Insights." IEEE Transactions on Systems, Man, and Cybernetics. 17(5):741-752.

Morgan, M.G., M. Henrion, and S.C. Morris (1980), "Expert Judgment for Policy Analysis." Brookhaven National Laboratory. BNL 51358.

Morgan, M.G., and M. Henrion (1990), Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, Cambridge University Press, New York.

National Council on Radiation Protection and Measurement, 1996, A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination, NCRP Commentary, Number 14, Bethesda, MD, 10 May 1996.

Price, P.S., S.H. Su, J.R. Harrington, and R.E. Keenan, 1996, Uncertainty and Variability in Indirect Exposure Assessments: An Analysis of Exposure to Tetrachlorodibenzo-p-dioxin from a Beef Consumption Pathway, Risk Analysis, Volume 16, Number 2, pp 263 - 277.

Rai, S.N., Krewksi, D., and Bartlett, S. 1996. "A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment," *Human and Ecological Risk Assessment*, **2**, 4, 972-989.

Spetzler, C.S., and Stael von Holstein (1975). "Probability Encoding in Decision Analysis." Management Science, 22(3).

US EPA (1996), Summary Report for the Workshop on Monte Carlo Analysis, EPA/630/R-96/010, September 1996.

Appendix A

Comments on "Extrapolation of Uncertainty of ASPEN Results (Revised)."

Prepared by:

H. Christopher Frey, Ph.D.
Assistant Professor
Department of Civil Engineering
North Carolina State University
Raleigh, NC 27695-7908

Prepared for:

US Environmental Protection Agency

When making estimates of uncertainty, it is important to clearly define the geographic area and averaging time. It is apparent that the averaging time used in the study is one year. However, it is less clear what the geographic area is in each case. The variation in P/O ratios is for an annual average, but for what geographic area? For example, are all ratios based upon a 50 km radius modeling area from a census tract centroid? \Presumably, the emission sources are site-specific, and are comprised of all emission sources of the HAP in question within a 50 km radius of the census tract centroid in question. It would be very helpful to have some diagrams, such as maps, that illustrate the geographic and spatial aspects of the modeling, with examples for both small and large (in land area) census tracts to illustrate various situations regarding location of emissions sources versus locations of receptors.

How source-specific are the emissions estimates? Potentially important sources of uncertainty in the emissions estimates include emission rates, source locations, stack parameters, and omissions of some sources. To what extent are surrogate emissions data used?

Why were census tract centroids not used as receptors in the air quality modeling?

The use of Gaussian plume modeling limits the assessment of air quality impacts to a distance of no more than 50 km from each emission source. It is indicated that effects due to long range transport are assumed to be accounted for in the background ambient air concentration estimates. However, medium or long range transport phenomena may not lead to a uniform background concentration throughout the entire U.S. Furthermore, if long-range transport due to U.S. emissions is treated as part of the background, then it will be difficult in the future to evaluate the benefits of emissions reductions with respect to long range transport. Even though long range transport may result in very low incremental air quality concentrations, it is possible that it may still result in significant population risks if large populations are affected.

In the "bottom-up" uncertainty analysis, it would be important to include uncertainty in the emission rate. For some HAPs and emission sources, annual average emission rates are likely to be uncertain by perhaps an order of magnitude or more.

In the discussion of the bottom-up approach, it is mentioned that building downwash was considered. This seems like a highly localized consideration that is not consistent with the objective of estimating average outdoor concentrations for a census tract, unless the geographic extent of the particular census tract is rather small.

In comparing the ASPEN inventory with the National Toxics Inventory (NTI), it would help to clarify whether the differences between the two were random or systematic. For example, it is stated that there was a difference of a factor of more than 3 for more than half of the HAPs. Were all of these underestimated when comparing ASPEN to NTI, or were there an approximately similar portion of underestimates and overestimates? A graphic providing a cumulative distribution function of the ratio of the estimates over all of the HAPs compared would be helpful.

It is not clear why there should be "difficulty in directly estimating the uncertainty of emissions estimates". There are methods for estimating uncertainty even in situations in which there are relatively few data (e.g., Frey and Rhodes, 1996).

On page 3 it is stated that "CO is expected to behave similarly to gaseous HAPs with very low reactive decay rates." However, what is not stated is the representativeness of this assumption. In particular, for which specific HAPs is this assumption considered to be valid? Clearly, this assumption is not correct for reactive HAPs or for particulate matter (PM). The latter suggests that, in addition to using CO as a basis for comparison, PM or PM_{10} should be considered in addition.

The basis for selecting and dealing with monitoring data is somewhat problematic. The focus on 1990 would appear to pose a substantial difficulty because of the relative lack of HAP monitoring stations at that time. The selection criteria also appear to be rather stringent. It would be useful to know how many monitoring stations were excluded from consideration because they did not meet the requirements for measurements over 10 continuous months or no more than 10 percent of values below the detection limit. For example, how many stations had data for 8 or 9 continuous months, or for 10 or 11 months but not continuously? What is considered to be "continuous"? How many stations had 15, 20, or 30 percent non-detect values that might otherwise have been considered as an acceptable station?

The treatment of nondetects is problematic. It is stated that "if a substantial fraction of the data are below the MDL, specifying values for them requires application of assumptions that may significantly influence the estimate of the annual average concentration." While this may be true, it is also possible to do bounding analyses to develop a maximum range of possible values for the annual average concentration (i.e. by comparing situations in which all values

below the MDL are assumed to be zero with one in which all such values are assumed to be the same as the MDL). Whether or not this maximum range of uncertainty affects any conclusions about comparisons of predicted to "observed" ambient concentrations can then be evaluated. For example, if the predicted values are low regardless of the range of uncertainty in the "observed" annual average, then it is possible that there are errors in emissions estimation or dispersion modeling. On the other hand, if the predicted value is within the range of uncertainty of the "observed" annual average, than it may be important to develop improved monitoring methods with lower MDLs in order to improve future comparisons. Thus, the comparison of predicted values to "observed" annual averages, even in cases with a large proportion of nondetected values, may still be useful.

Furthermore, the approach taken for handling nondetected values is not a particularly satisfactory one (i.e. assuming one-half of the detection limit for all data below the detection limit). An alternative to the bounding cases described in the previous paragraph would be to develop more plausible estimates of the annual average by fitting probability distributions to the observed data and extrapolating into the non-detect range. For example, maximum likelihood estimation (MLE) can be used to fit a parametric distribution to a data set that contains non-detected values. Currently, with one of my graduate students I am performing numerical experiments with this approach. We have evaluated, for example, Normal, Lognormal, Gamma, Weibull, and Beta distributions fitted to data sets of sample size 20 and 50 with varying proportions of non-detected values (e.g., 5, 10, 15, and 20 non-detected values in the case of a data set of size 50). Typically, there is little variation in the fit for the values of data set that are above the detection limit, and reasonable consistency of the fit for values below the detection limit. By fitting a distribution to the data, one can then make an estimate of the mean value. We are in the process of developing and demonstrating an approach for characterizing uncertainty in the fit of the distribution. This will enable calculation of a probability distribution for uncertainty in the mean value.

To the extent that additional data sets might become available by making reasonable relaxations to the selection criteria (e.g., accepting data sets where 20-40 percent of the values were below the MDL instead of only 10 percent), it would be worthwhile to employ more sophisticated methods for making extrapolations for non-detected data and for evaluating uncertainty in the resulting estimate of the annual average pollutant concentration.

A potentially significant issue that is not addressed is the measurement errors for the monitoring data. If the measurement errors are small, then any discrepancies between the predicted and observed values might be attributable to errors in emissions estimation and/or dispersion modeling. However, if measurement errors are large, then the distribution of the ration of predicted to observed values may be merely due to measurement errors. Thus, we are interested in knowing how large a discrepancy must exist between the predicted and observed values before we can attribute it to a systematic error in the modeling approach, as opposed to either systematic and/or random error in the measurement methods. Furthermore, to the extent that different measurement methods were used as a basis for emissions estimation and

for ambient air quality monitoring, there is a possibility of different systematic errors in each case that could complicate comparisons.

On page 6 it is not clear how background concentrations are accounted for in the approach. Are the "observed" values based upon subtracting background estimates from the annual average measurement at the monitoring site? Or is it assumed that the background concentration is included in the "predicted" value, as hinted at on page 4? How does the background concentration compare with typical estimates of concentrations attributable to emissions from census tracts? Can anything be said, even qualitatively, about the potential uncertainties in background levels in comparison to the uncertainties in concentrations attributable to quantified emissions and short-range transport? For example, if the estimated concentration in a census tract is 10 times greater than the estimated and assumed national background concentration, is it possible that background concentration might nonetheless be the dominant source of uncertainty at that particular location?

On page 6 is it stated that it is assumed that estimated emissions of CO are the same as the actual emissions of CO. In other words, CO emissions estimates are assumed to be precise and accurate. Based upon this assumption, if the ratios of predicted and observed CO differed from one, the explanation would be based upon failure to consider actual dispersion conditions. However, to the extent that the CO emissions estimates are either imprecise and/or inaccurate, then descrepancies between observed and predicted CO concentrations could be due to errors in emissions.

I have many comments regarding the treatment of mobile sources and in particular the discussion of biases in CO emissions and the use of the Mobile5a model. These comments are based upon my extensive experience in probabilistic analysis of the Mobile5a model (e.g., Frey and Kini, 1997). I have also served as a peer reviewer for a recent Office of Mobile Sources document regarding key assumptions underlying the development of Mobile6.

It is stated that estimates of CO emissions "are expected to be reasonably accurate, with some probability of being underestimated by less than 25%". There is some confusion on what this means. In Appendix A it is stated that "approximately 25 percent of the light duty auto CO emissions was due to off-cycle vehicle operation." If this is assumed to be true, then the implication is that we would have to increase the CO emission estimates by a factor of 1.33.

In Appendix A there appears to be some misunderstanding of the Mobile5 model. The speeds that are entered into the model represent average speeds for a driving cycle. Thus, even if the highest input speed was 58.4 mph, this does not mean that more extreme speeds were not considered in the emission factor. Consider the example of the Federal Test Procedure (FTP). The FTP has an average speed of 19.6 mph, but the instantaneous vehicle speeds during the test vary from 0 to 57 mph. There are driving cycles with higher average speeds, such as the Highway Fuel Economy Test (HFET) and several California Air Resources Board (ARB) cycles, which also have higher peak speeds. Nonetheless, it is true that these cycles

underestimate not so much high speeds as they understimate high accelerations or combinations of speed and acceleration associated with high engine loads.

It is not at all reasonable to expect to produce emission factors for a particular road, as suggested in Appendix A. The Mobile5 model can only be used in a credible fashion for making average predictions for substantially large vehicle fleets and for entire trips.

The discussion of comparison of tunnel studies and the Mobile model is incorrect. In particular, the statement "tunnel studies tend to represent relatively steady state driving conditions with "warmed up" vehicle, which are conditions where one might expect MOBILE to perform reasonably well in relation to observations" is not accurate. Mobile emission factors are not based upon steady state driving; they are based upon driving cycles which in turn are based upon dynamic variations in speed and acceleration. For those tunnels that have free-flow, congestion-free traffic conditions, one would expect a bias in the comparison with Mobile5, because Mobile5 is not able to represent such situations. In fact, the comparisons presented as an example on page 88 appear to be quite reasonable, assuming that traffic in the tunnel was moving more smoothly than the simulated vehicle movement assumed in the driving cycles that underlie Mobile5. Furthermore, it is incorrect to compare a segment or link-based emission estimate for a tunnel with a trip-based estimate from Mobile5. The Mobile5 model cannot be used to make an estimate of emissions over a short segment of one roadway. This is because the driving cycles are based upon an entire trip, from start to finish. A trip may occur over a variety of roadway facilities, not just the facility type represented by the tunnel. For these reasons, one expects to find biases in the comparison of tunnel studies to the Mobile5a model. The widespread misinterpretation of the meaning of these comparisons can typically be traced to lack of knowledge regarding the basis for the Mobile5a model. This is understandable, given the relative lack of documentation of that model. To EPA's credit, a significant effort is being made to develop a more credible approach to emissions estimation in the forthcoming Mobile6 model, to submit key assumptions of the new model to peer review, and to more fully document the new model.

Frey and Kini (1997) have done a probabilistic analysis of the Mobile5a model. This analysis involved reanalyzing data sets pertaining to light duty gasoline vehicle emissions for selected technology groups. One of the key findings was that the precision of the model predictions is typically no better than plus or minus 25 percent for a 90 percent probability range. Furthermore, there are some biases in the model predictions due to the mathematical formulation of the model. Not accounted for in that study are additional biases and imprecision due to non-representativeness of the driving cycles with respect to on-road driving.

It is unclear, on page 6 and Appendix A, whether the CO emission inventory was adjusted to account for potential biases. For example, was the on-road mobile sources emission inventory multiplied by a factor of 1.33 to account for off-cycle events?

On page 7 it is mentioned that the closest CO monitors typically ranged from 0 to 413 km from each HAP monitor. Since the Gaussian plume model is not considered to be valid for predictions beyond 50 km, it appears to be problematic to make comparisons among monitoring stations as far apart as 413 km. What portion of HAP monitors were more than, say, 50 km distant from the nearest CO monitor? Are both the CO and HAP monitor considered to be at location "X" even if they are in reality more than 50 km apart? If the purpose of normalizing HAP comparisons to CO comparisons at the same site is to screen out dispersion as a factor in differences between observed and predicted concentrations, it would appear to be self-defeating to assume that dispersion conditions at a CO monitor several hundred kilometers away would be representative of conditions at the HAP monitor. In any event, since the Guassian plume model should not be extrapolated, it would appear necessary to use CO monitoring data as a basis for adjustments only if it is within 50 km of the HAP monitor.

The two equations on Page 8 appear to be the same; thus, one must be in error. Furthermore, it would be extremely helpful to provide numerical examples to demonstrate how these equations are used.

The material presented at the bottom of page 8 is poorly defined and rather confusing. It is not clear why all of this information is presented. Any time an equation is presented all of the variables should be clearly defined. Furthermore, it is usually helpful to give a numerical example. The basis for the five algorithms used in SAS is not given; thus, it is unclear what the various relationships are intended to represent or what their potential advantages or disadvantages are. The equation given at the top of page 9 is not well motivated. Why was this selected? What is the interpretation of it? How is it used (what does "b" represent? What does "g" represent? etc.). Provide a numerical example of how to use it. It seems likely that some of the material on the bottom of page 8 is not needed. All that is needed is to present the approach used and enough information to justify it.

The "1 sample Wilcoxon signed rank test" should be explained, and a reference should be cited for it.

Some more critical attention is needed regarding the interpretation of the "uncertainty intervals." These intervals are based upon variability in the predicted-to-observed ratios ("P/O ratios") from one location to another for a given HAP. As such, these are not "uncertainty" intervals. The interpretation of these in terms of "uncertainty" is based upon an assumption that the variability in the P/O ratio is either unexplainable or is as yet unexplained (in a quantitative sense). After stating this assumption, then it would be possible to refer to these as uncertainty intervals. It should be clearly stated that these intervals are based upon 90 percent probability ranges, which is perhaps not the most standard probability interval to use (95 percent might be a more common one).

In the discussion of formaldehyde, it might be helpful to use the terms "primary" and "secondary" pollutant, to clarify that formaldehyde is emitted directly in some cases and is formed in the atmosphere in other cases as the result of chemical reactions in the atmosphere. It should be pointed out that formaldehyde is also reactive, in that it has a relatively short lifetime in the atmosphere compared to CO. Hence, it is not clear that it is useful to adjust the P/O ratio based upon comparison with CO.

It is not evident that formaldehyde has a higher level of uncertainty than other compounds, as stated on the bottom of page 9. There are 5 other compounds with an equal or higher level of uncertainty than that for formaldehyde.

For tetrachloroethylene P/O ratios, it appears that the range of variability in the ratios is lower for California sites than for non-California sites. More thorough interpretation would be helpful. Is this because emission inventories in California might be more complete and more accurate? Or is it due to less variation in disperson characteristics? Or some combination of the two?

On page 11, in the paragraph just after the middle of the page, it is stated that "the procedure used to account for uncertainty due to dispersion can change the expected value of the "true" concentration substantially." A change in the uncertainty range from a factor of 7.5 to a factor of 5 is not particularly "substantial", nor is a change in the percentage of values outside of the interval from 6 percent to 10 percent. Thus, the word "substantially" does not appear to be appropriate here.

On page 12, last paragraph before the summary and recommendations, it would be useful to provide more interpretation of the data given in Tables 11-13 and Figures 19-24.

I would also like to see some empirical cumulative distribution functions (ECDFs) for the variation in the P/O ratios, at least for some selected cases. Similarly, I would like to see the ECDFs for the P/O ratios proposed for use in adjusting for HAPs without monitoring data.

As noted previously, but pertinent to the discussion on the bottom of page 13, it would be useful to evaluate the use of PM monitoring data as a basis for making dispersion and deposition adjustments for HAPs that are associated with PM. Failure to properly model or adjust particulate-HAPs ambient air quality predictions could lead to substantial overestimation of these concentrations. While a biased overestimation may be useful for a screening analysis, it could lead to substantial problems of public perception and misallocation of environmental protection resources if the results are misinterpreted.

The recommendation on page 15 regarding the application of the dispersion adjustment approach to particulate HAPs should be stated as an interim recommendation, along with a recommendation that the sensitivity of this assumption should be explored in future work. It is not credible to use this approach for highly reactive gaseous pollutants or for particulate

HAPs, and any use of this approach in the short term should be viewed only as a stopgap measure to develop bounding estimates pending development of a better approach.

Overall, the use of variability in P/O ratios as a means for gaining insight into uncertainty in HAP emissions and dispersion predictions is useful, but subject to many limitations as described in the report. It is apparent that there is a large range of uncertainty, which is not surprising. It would be useful to place this uncertainty in perspective by doing some "model" bottoms-up analyses (which may already have been done). It does not appear that any attention has been given to uncertainty in emissions rates or uncertainties due to measurement errors of both emissions and ambient concentrations. Such measurements would represent constraints on the lower limit on the range of uncertainty that could be expected in a study such as this. Thus, it would be useful to quantify these uncertainties for comparison with the P/O ratios.

It seems unlikely that the model could be expected to make accurate predictions at a census-tract level given the current state of information, depending upon the geographic extent of the census tract, among other factors. The basis for reporting results should be carefully considered. It is probably not unreasonable to report results at some higher level of geographic aggregation, such as county, metropolitan area, or state.

References:

Frey, H.C., and Kini, M.D. (1997). Probabilistic Modeling of Mobile Source Emissions. Report Prepared for Center for Transportation and the Environment by North Carolina State University. (contact author for complete citation).

Frey, H.C., and D.S. Rhodes (1996), "Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example," *Human and Ecological Risk Assessment*, 2(4) (December 1996)

Appendix B

U.S. EPA Risk Assessment Forum Workshop on "Selecting Input Distributions for Probabilistic Analysis" held April 21-22, 1998, New York City

Summary
(DRAFT)
Prepared by the Chair
H. Christopher Frey
North Carolina State University
June 1998

The workshop was comprised of five major sessions, in which three were devoted to the issue of representativeness and two were devoted to issues regarding parametric versus empirical distributions and goodness-of-fit. Each session began with a trigger question. For the three sessions on representativeness, there was discussion in a plenary setting, as well as discussions within four break-out groups. For the two session regarding selection of parametric versus empirical distributions and the use of goodness-of-fit tests, the discussions were conducted in plenary sessions.

Representativeness

The first session was devoted to three main questions, based upon the portion of the workshop charge requesting feedback on the representativeness issue paper. After some general discussion, three trigger questions were formulated and posed to the group. These were:

1. What information is required to fully specify a problem definition?
2. What constitutes (lack of) representativeness?
3. What considerations should be included in, added to, or excluded from the checklists?

The group was then divided into four break-out groups, each of which addressed all three of these questions. Each group was asked to use an approach known as "brainwriting." Brainwriting is intended to be a silent activity in which each member of a group at any given time puts thoughts down on paper in response to a trigger question. After completing an idea, a group member exchanges papers with another group member. Typically, upon reading what others have written, new ideas are generated and written down. Thus, each person has a chance to read and respond to what others have written. Advantages of brainwriting are that all panelists can be generating ideas simultaneously, there is less of a problem with domination of the discussion by just a few people, and a written record is produced as part of the process. A disadvantage is that there is less "interaction" with the entire panel. After the brainwriting activity was completed, a representative of each panel reported the main ideas to the entire group.

The panel generally agreed that before addressing the issue of representativeness, it is necessary to have a clear problem definition. Therefore, there was considerable discussion of what factors must be considered to ensure a complete problem definition. The most general criteria for a good problem definition, to which the group gave general assent, is to specify the "who, what, when, where, why, and how". The "who" addresses what population is of interest. "Where" addresses the spatial extent of the assessment. "When" addresses the temporal extent of the assessment. "What" relates to the specific chemicals and health effects of concern. "Why" and "how" may help clarify the previous matters. For example, it is helpful to known that exposures occur because of a particular behavior (e.g., fish consumption) when attempting to define an exposed population and the spatial and temporal characteristics of the problem. Knowledge of "why" and "how" is also useful later for proposing mitigation or prevention strategies. The group in general agreed upon these principles for a problem definition, as well as the more specific suggestions detailed in Section 4.1.1.

In regard to the second trigger question, the group generally agreed that "representativeness" is context-specific. Furthermore, there was a general trend toward finding other terminology instead of using the term "representativeness". In particular, many panelists concurred that an objective in an assessment is to make sure that it is "useful and informative" or "adequate" for the purpose at hand. The adequacy of an assessment may be evaluated with respect to considerations such as "allowable error" as well as practical matters such as the ability to make measurements that are reasonably free of major errors or to reasonably interpret information from other sources that is used as an input to an assessment. Adequacy may be quantified, in principle, in terms of the precision and accuracy of model inputs and model outputs. There was some discussion of how the distinction between variability and uncertainty relates to assessment of adequacy. For example, one may wish to have accurate predictions of exposures for more than one percentile of the population, reflecting variability. For any given percentile of the population, however, there may be uncertainty in the predictions of exposures. Some panelists pointed out, that, because often it is not possible to fully validate many exposure predictions or to obtain input information completely free of error or uncertainty, there is an inherently subjective element in assessing adequacy. The stringency of the requirement for adequacy will depend upon the purpose of the assessment. It was noted, for example, that it may typically be easier to adequately define mean values of exposure than upper percentile values of exposure. Adequacy is also a function of the level of detail of an assessment: the requirements for adequacy of an initial, screening level calculation will typically be less rigorous than those for a more detailed analysis.

Regarding the third trigger question, the panel was generally complimentary of the proposed checklists in the representativeness issue paper. Of course, the panel had many suggestions for improvements in the checklists. Some of the broader concerns were about how to make the checklists context-specific, since the degree of usefulness of information depends on both the quality of the information and upon the purpose of the assessment. Some of the specific suggestions included use of flowcharts rather than lists, avoiding overlap among the flowcharts or lists, development of an interactive web-based flowchart that would be flexible

and context-specific, and clarification of terms used in the issue paper (e.g., "external" versus "internal" distinction). The panel also suggested that the checklists or flowcharts should encourage additional data collection where appropriate, and should promote a "value of information" approach to help prioritize additional data collection. Further discussion of the panel's comments is given in Section 4.1.3.

Sensitivity Analysis

The second session was devoted to issues encapsulated in the following trigger questions:

How can one do sensitivity analysis to evaluate the implications of non-representativeness? In other words, how do we assess the importance of non-representativeness?

The panel was asked to consider data, models and methods in answering these questions. Furthermore, the panel was asked to keep in mind that the charge requested recommendations for immediate, short-term, and long-term studies or activities that could be done to provide methods or examples for answering these questions.

There were a variety of answers to these questions. A number of panelists shared the view that non-representativeness may not be important in many assessments. Specifically, they argued that many assessments and decisions consider a range of scenarios and populations. Furthermore, populations and exposure scenarios typically change over time, so that if one were to focus on making an assessment "representative" for one point in time or space, it could fail to be representative at other points in time or space or even for the original population of interest as individuals enter, leave, or change within the exposed population. Here again the notion of adequacy, rather than representativeness, was of concern to the panel. The panel also reiterated that representativeness is context-specific. Furthermore, there was some discussion of situations in which data are collected for "blue chip" distributions that are not specific to any particular decision.

The panel did recommend that, in situations where there may be a lack of adequacy of model predictions based upon available information, the sensitivity of decisions should be evaluated under a range of plausible adjustments to the input assumptions. It was suggested that there may be multiple tiers of analyses, each with a corresponding degree of effort and rigor regarding sensitivity analyses. In a "first tier" analysis, the use of bounding estimates may be sufficient to establish sensitivity of model predictions with respect to one or more model outputs, without need for doing a probabilistic analysis. After a preliminary identification of sensitive model inputs, the next step would typically be to develop a probability distribution to represent a plausible range of outcomes for each of the sensitive inputs. Key questions to be considered are whether to attempt to make adjustments to improve the adequacy or representativeness of the assumptions and/or whether to collect additional data to improve the characterization of the input assumptions.

One potentially helpful criteria for deciding whether data are adequate is to try to answer the question: "are the data good enough to replace an assumption?" If not, then additional data collection is likely to be needed. One would need to assess whether the needed data can be collected. A "value of information" approach can be useful in prioritizing data collection and in determining when sufficient data have been collected.

There was some discussion of sensitivity analysis of uncertainty versus sensitivity analysis of variability. The panel generally agreed that sensitivity analysis to identify key sources of uncertainty is a useful and appropriate thing to do. There was disagreement among the panelists regarding the meaning of identifying key sources of variability. One panelist argued that identifying key sources of variability is not useful, because variability is irreducible. However, knowledge of key sources of variability can be useful in identifying key characteristics of highly exposed subpopulations or in formulating prevention or mitigation measures.

In the present, there are many methods that already exist for doing sensitivity analysis, including running models for alternative scenarios and input assumptions and the use of regression or statistical methods to identify the most sensitive input distributions in a probabilistic analysis. In the short to long term, it was suggested that some efforts be devoted to the development of "blue chip" distributions for quantities that are widely used in many exposure assessments (e.g., intake rates of various foods). It was also suggested that new methods for sensitivity analysis might be obtained from other fields, with specific examples based upon classification schemes, time series, and "g-estimation".

Making Adjustments to Improve Representation

In the third session, the panel responded to the following trigger question:

How can one make adjustments from the sample to better represent the population of interest?

The panel was asked to consider "population", spatial, and temporal characteristics when considering issues of representativeness and methods for making adjustments. The panel was asked to provide input regarding exemplary methods and information sources that are available now to help in making such adjustments, as well as to consider short-term and long-term research needs.

The panel clarified some of the terminology that was used in the issue paper and in the panel's discussion. The term "population" was defined as referring to "an identifiable group of people." The panel noted that often one has a sample of data from a "surrogate population", which is not identical to the "target population" of interest in a particular exposure assessment. The panel noted that there is a difference between "analysis" of actual data pertaining to the target population, versus "extrapolation" of information from data for a

surrogate population to make inferences regarding a target population. It was noted that extrapolation always "introduces" uncertainty.

On the temporal dimension, the panel noted that one potential problem occurs when data are collected at one point in time and used in an assessment aimed at a different point in time because of shifts in the characteristics of populations between the two time periods.

Reweighting of data was one approach that was mentioned in the plenary discussion. There was a discussion of "general" versus mechanistic approaches for making adjustments. The distinction here was that "general" approaches might be statistical, mathematical, or empirical in their foundations (e.g., regression analysis) whereas mechanistic approaches would rely on theory specific to a particular problem area (e.g., a physical, biological, or chemical model). It was noted that temporal and spatial issues are often problem-specific, which makes it difficult to recommend universal approaches for making adjustments. The panel generally agreed that it is desirable to include or state uncertainties associated with extrapolations. Several panelists strongly expressed the view that "it is okay to state what you don't know," and there was no disagreement on this point.

The panel recommended that the basis for making any adjustments to assumptions regarding populations should be predicated upon stakeholder input and the examination of covariates. The panel noted that methods for analyzing spatial and temporal aspects exist, if data exists. Of course, a common problem is a scarcity of data and a subsequent reliance on surrogate information. For assessment of spatial variations, methods such as kreiging (sp?) and random fields were commonly suggested. For assessment of temporal variations, time series methods were suggested.

There was a lively discussion regarding whether adjustments should be "conservative". Some panelists initially argued that, in order to protect public health, any adjustments to input assumptions should tend to be biased in a conservative manner (so as not to make an error of understating a health risk, but with some non-zero probability of making an error of overstating a particular risk). After some additional discussion, it appeared that the panel was in agreement that one should strive primarily for accuracy, and that ideally any adjustments that introduce "conservatism" should be left to decision makers. It was pointed out that invariably many judgments go into the development of input assumptions for an analysis, and that these judgments in reality often introduce some conservatism. Several pointed out that "conservatism" can entail significant costs if it results in over-control or misidentification of important risks. Thus, conservatism in individual assessments may not be optimal or even conservative in a broader sense, if some sources of risk are not addressed because others receive undue attention. Therefore, the overall recommendation of the panel regarding this issue is to strive for accuracy rather than conservatism, leaving the latter as an explicit policy issue for decision makers to introduce, although it is clear that individual panelists had somewhat differing views.

The panel's recommendations regarding measures that can be taken now include the use of stratification to try to reduce variability and correlation among inputs in an assessment, brainstorming to generate ideas regarding possible adjustments that might be made to input assumptions, and stakeholder input for much the same purpose, as well as to make sure that no significant pathways or scenarios have been overlooked. It was agreed that "plausible extrapolations' are reasonable when making adjustments to improve representativeness or adequacy. What is "plausible" will be context-specific.

In the short term, the panel recommends that the following activities be conducted:

¥ Numerical Experiments. Numerical experiments can be used to test existing and new methods for making adjustments based upon factors such as averaging times or averaging areas. For example, the precision and accuracy of the Duan-Wallace model for making adjustments from one averaging time to another can be evaluated under a variety of conditions via numerical experiments.

¥ Workshop on Adjustment Methods. The panel agreed in general that there are many potentially useful methods for analysis and adjustment, but that many of these are to be found in fields outside of the risk analysis community. Therefore, it would be useful to convene a panel of experts from other fields for the purpose of cross-disciplinary exchange of information regarding methods applicable to risk analysis problems. For example, it was suggested that geostatistical methods should be investigated.

¥ Put Data on the Web. There was a fervent plea from at least one panelist that data for "blue chip" and other commonly used distributions should be placed on the web, to facilitate dissemination and analysis of such data. A common concern is that often times data are reported in summary form, which makes it difficult to analyze the data (e.g., to fit distributions). Thus, the recommendation includes the placement of actual data points, and not just summary data, on publicly accessible web sites.

¥ Suggestions on How to Choose A Method. Although the panel felt it was unrealistic to provide recommendations regarding specific methods for making adjustments, because of the potentially large number of methods and the need for input from people in other fields, the panel did suggest that it would be possible to create a set of criteria regarding desirable features for such methods that could help an analyst when making choices among many options.

In the longer term, the panel recommends that efforts be directed at more data collection, such as improved national or regional surveys, to better capture variability as a function of different populations, locations, and averaging times. Along these lines, specific studies could be focused on the development or refinement of a select set of "blue chip" distributions, as well as targeted at updating or extending existing data sets to improve their flexibility for use in assessments of various populations, locations, and averaging times. The panel also noted that

because populations, pathways, and scenarios change over time, there will be a continuing need to improve existing data sets.

Empirical and Parametric Distribution Functions

In the fourth session, the panel began to address the second main set of issues as given in the charge. The trigger question used to start the discussion was:

What are the primary considerations in choosing between the use of parametric distribution functions (PDFs) and Empirical Distribution Functions (EDFs)?

The panel was asked to consider the advantages of using one versus the other, whether the choice is merely a matter of preference, whether one is preferred, and whether there are cases when neither should be used.

The initial discussion involved clarification of the difference between the terms EDF and "bootstrap". Bootstrap simulation is a general technique for estimating confidence intervals and characterizing sampling distributions for statistics, as described by Efron and Tibshirani (1993). An EDF can be described as a stepwise cumulative distribution function or as a probability density function in which each data point is assigned an equal probability. Non-parametric bootstrap can be used to quantify sampling distributions or confidence intervals for statistics based upon the EDF, such as percentiles or moments. Parametric bootstrap methods can be used to quantify sampling distributions or confidence intervals for statistics based upon PDFs. Bootstrap methods are often referred to also as "resampling" methods. However, "bootstrap" and EDF are not the same thing.

The panel generally agreed that the choice of EDF vs. PDF is usually a matter of preference, and also expressed the general opinion that there should be no rigid guidance requiring the use of one or the other in any particular situation. The panel briefly addressed the notion of consistency. While consistency in the use of a particular method (e.g., EDF or PDF, in this case) may offer benefits in terms of simplifying analyses and helping decision makers, there was a concern that any strict enforcement of consistency will inhibit the development of new methods or the acquisition of new data and may also lead to compromises from better approaches that are context-specific. Here again it is important to point out that the panel explicitly chose not to recommend the use of either EDF or PDF as a single preferred approach, but rather to recommend that this choice be left to the discretion of analysts on a case-by-case basis. For example, it could be reasonable for an analyst to include EDFs for some inputs and PDFs for others even within the same analysis.

Some panelists gave examples of situations in which they might personally prefer to use an EDF, such as: (a) when there are a large number of data points (e.g., 12,000); (b) access to high

speed data storage and retrieval systems; (c) when there is no theoretical basis for selecting a PDF; and/or (d) when one has an "ideal" perfect sample. There was some discussion of preference for use of EDFs in "data rich" situations rather than "data poor" situations. However, it was noted that "data poor" is context-specific. For example, a data set may be adequate for estimating the 90th percentile, but not the 99th percentile. Therefore, one may be "data rich" in the former case and "data poor" in the latter case with the same data set.

Some panelists also gave examples of when they would personally prefer to use PDFs. A potential limitation of conventional EDFs is that they are restricted to the range of observed data. In contrast, PDFs typically provide estimates of "tails" of the distribution beyond the range of observed data, which may have intuitive or theoretical appeal. PDFs are also preferred by some because they provide a compact representation of data and can provide insight into generalizable features of a data set. Thus, in contrast to the proponent of the use of an EDF for a data set of 12,000, another panelist suggested it would be easier to summarize the data with a PDF, as long as the fit was reasonable. At least one panelist suggested that a PDF may be easier to defend in a legal setting, although there was no consensus on this point.

For both EDFs and PDFs the issue of extrapolation beyond the range of observed data received considerable discussion. One panelist stated that the "further we go out in the tails, the less we know," to which another panelist responded "when we go beyond the data, we know nothing." As a rebuttal, a third panelist asked "do we really know nothing beyond the maximum data point?" and suggested that analogies with similar situations may provide a basis for judgments regarding extrapolation beyond the observed data. Overall, most or all of the panelists appeared to be supportive of some approach to extrapolation beyond observed data, regardless of whether one prefers an EDF or PDF. Some argued that one has more control over extrapolations with EDFs, because there are a variety of functional forms that can be appended to create a "tail" beyond the range of observed data. Examples of these are described in the issue paper. Others argued that when there is a theoretical basis for selecting a PDF, then there is also some theoretical basis for extrapolating beyond the observed data. It was pointed out that one should not always focus on the "upper" tail; sometimes the lower tail of a model input may lead to extreme values of a model output (e.g., such as when an input appears in a denominator).

There was some discussion of situations in which neither an EDF or PDF may be particularly desirable. One suggestion was that there may be situations in which explicit enumeration of all combinations of observed data values for all model inputs, as opposed to a probabilistic resampling scheme, may be desired. Such an approach can help, for example, in tracing combinations of input values that produce extreme values in model outputs. One panelist suggested that neither EDFs nor PDFs are useful when there must be large extrapolations into the tails of the distributions.

A question that the panel chose to address was "how much information do we lose in the tails of a model output by not knowing the tails of the model inputs?" One comment was that it may not be necessary to accurately characterize the tails of all model inputs because the tails (or extreme values) of model outputs may depend on a variety of other combinations of model input values. Thus, it is possible that even if no effort is made to extrapolate beyond the range of observed data in model inputs, one may still predict extreme values in the model outputs. The use of scenario analysis was suggested as an alternative or supplement to probabilistic analysis in situations in which either a particular input cannot reasonably be assigned a probability distribution or when it may be difficult to estimate the tails of an important input distribution. In the latter case, alternative upper bounds on the distribution, or alternative assumptions regarding extrapolation to the tails, should be considered as scenarios.

Uncertainty in EDFs and PDFs was discussed. Techniques for estimating uncertainties in the statistics (e.g., percentiles) of various distributions, such as bootstrap simulation, are available. An example was presented, for a data set comprised of six measurements, illustrating how the uncertainty in the fit of a parametric distribution was greatest at the tails. It was pointed out when considering alternative PDFs (e.g., Lognormal vs. Gamma) the range of uncertainty in the upper percentiles of the alternative distributions will typically overlap; therefore, apparent differences in the fit of the tails may not be particularly significant from a statistical perspective. Such insights are obtained from an explicit approach to distinguishing between variability and uncertainty in a "two-dimensional" probabilistic framework.

The panel discussed whether mixture distributions are useful. Some panelists were clearly proponents of using mixture distributions. A few panelists offered some cautions that it can be difficult to know when to properly employ mixtures. One example mentioned was for radon concentrations. One panelist mentioned in passing that radon concentrations had been addressed in a particular assessment assuming a lognormal distribution. Another responded that the concentration may more appropriately be described as a mixture of normal distributions. There was no firm consensus on whether it is better to use a mixture of distributions as opposed to a "generalized" distribution that can take on many arbitrary shapes. Those who expressed opinions tended to prefer the use of mixtures since they could offer more insight about processes that produced the data.

Truncation of the tails of a PDF was discussed. Most panelists seemed to view this as a last resort fraught with imperfections. The need for truncation may be the result of an inappropriate selection of a PDF. For example, one panelist asked "if you truncate a Lognormal, does this invalidate your justification of the Lognormal?" It was suggested that alternative PDFs (perhaps ones that are less "tail-heavy") be explored as an alternative. Some suggested that truncation is often unnecessary. Depending upon the probability mass of the portion of the distribution that is considered for truncation, the probability of sampling an extreme value beyond a plausible upper bound may be so low that it does not occur in a

typical Monte Carlo simulation of only a few thousand iterations. Even if an unrealistic value is sampled for one input, it may not produce an extreme value in the model output. If one does truncate a distribution, it can potentially affect the mean and other moments of the distribution. Thus, one panelist summarized the issue of truncation as "nitpicking" that potentially can lead to more problems than it solves.

Goodness-of-Fit

The fifth and final session of the workshop was devoted to the following trigger question:

On what basis should it be decided whether a data set is adequately fitted by a parametric distribution?

The premise of this session was the assumption that a decision had already been made to use a PDF instead of an EDF. While not all panelists were comfortable with this assumption, all agreed to base the subsequent discussion upon it.

The panel agreed unanimously that visualization of both the data and the fitted distribution is the most important approach for ascertaining the adequacy of fit. The panel in general seemed to share a view that conventional Goodness-of-Fit (GoF) tests have significant shortcomings, and that they should not be the only or perhaps even primary methods used for determining the adequacy of fit.

One panelist elaborated that any type of probability plot that allows one to transform data so that they can be compared to a straight line, representing a perfect fit, is extremely useful. The human eye is generally good at identifying discrepancies from the straight line perfect fit. Another panelist pointed out that visualization and visual inspection is routinely used in the medical community for evaluation of information such as x-rays and CAT-scans; thus, there is a credible basis for reliance on visualization as a means for evaluating models and data.

One of the potential problems with GoF tests is that they may be sensitive to imperfections in the fit that are not of serious concern to an analyst or decision maker. For example, if there are outliers at the low or middle portions of the distribution, a GoF test may suggest that a particular PDF should be rejected even though there is a good fit at the upper end of the distribution. In the absence of a visual inspection of the fit, the analyst may have no insight as to why a particular PDF was rejected by a GoF test.

The power of GoF tests was discussed. The panel in general seemed comfortable with the notion of overriding the results of a GoF test if what appeared to be a good fit, via visual inspection, was rejected by the test, especially for large data sets or when the imperfections are in portions of the distribution that are not of major concern to the analyst or decision maker. Some panelists shared stories of situations in which they have found that a particular GoF test would reject a distribution due to only a few "strange" data points in what otherwise appears

to be a plausible fit. It was noted that GoF tests become increasingly sensitive as the number of data points increases, so that even what appear to be small or negligible "blips" in a large data set are sufficient to lead to rejection of the fit. In contrast, for small data sets, GoF tests tend to be "weak" and may fail to reject a wide range of PDFs. One panelist expressed concern that any strict requirement for the use of GoF tests might reduce incentives for data collection, since it is relatively easy to avoid rejecting a PDF with few data.

The basis of GoF tests sparked some discussion. The "loss functions" assumed in many tests typically have to do with deviation of the fitted cumulative distribution function from the EDF for the data set. Other criteria are possible and in principal one could create any arbitrary GoF test. One panelist asked whether minimization of the loss function used in any particular GoF test might be used as a basis for choosing parameter values when fitting a distribution to the data. There was no specific objection, but it was pointed out that a degree-of-freedom correction would be needed. Furthermore, other methods, such as maximum likelihood estimation (MLE), have a stronger theoretical basis as a method for parameter estimation.

The panel discussed the role of the "significance level" and the "p-value" in GoF tests. One panelist stressed that the significance level should be determined in advance of evaluating GoF, and that it must be applied consistently in rejecting possible fits. Other panelists, however, suggested that the appropriate significance level would depend upon risk management objectives. One panelist suggested that it is useful to know the p-value of every fitted distribution, so that one may have an indication of how good or weak the fit may have been according to the particular GoF test.

APPENDIX A-4

Final Comments on the Draft Residual Risk Report to Congress
Science Advisory Board Residual Risk Subcommittee
Thomas Gentile
NYS Department of Environmental Conservation , Albany, NY
August 6, 1998

Charge element 1. **Within the context and scope of section 112 (f) (1) requirements has the Residual Risk Report to Congress properly interpreted and considered technical advice from previous reports, including: (1) the NRC's 1994 report *Science and Judgement in Risk Assessment* and (2) the 1977 report from the Commission on Risk Assessment and Risk Management (CRARM) in developing its risk assessment methodology residual risk strategy?**

Overall, Residual Risk Report to Congress (RTC) has considered the technical advice from the previous reports by acknowledging and discussing the practical acceptability of the various recommendations made by the NRC Committee and the Commission. The report succinctly describes how residual risk analyses for public health protection have been performed in the past and allows insight into how the Agency would like to proceed. However, the proper interpretation of the technical advice provided by the previous reports is difficult to make due to the general nature or open ended discussions about how the Agency will conduct a full residual risk assessment. A comprehensive discussion on the interpretation of the technical advice will have to wait until the risk assessment (RA) methodologies and risk management (RM) decision process described in the RTC are actually applied to various source categories by the Agency. However, the EPA acknowledged during their presentation that one of next steps would be the completion of the risk assessment methods (for determining non-cancer and ecological significance) and the presentation of case studies which cover all aspects of the application of residual risk assessment methods outlined in the RTC for SAB review in 1999. This should be noted in the RTC so Congress will not criticize the report for being deficient about the actual application of the RA and RM methods as discussed in the RTC.

An appendix in addition to the benzene decision, which provides a case-study on how the Agency or State or Local Air Toxics Agencies have conducted risk assessments on a specific source category (e.g. Municipal Waste Combustion Facilities) and the subsequent risk management decisions made by the governmental Agency about the significance of remaining risk would provide useful information to Congress. Another alternative would be to present the risk management guidelines used by the States in making permitting decisions about the significance of risk from HAP exposure.

Science and Judgement in Risk Assessment provided a set of several common themes which the Agency should address in the RTC: default options, data needs, validation, uncertainty, variability and aggregation, and four central themes made by the Committee on Risk Assessment of Hazardous Air Pollutants in their overall conclusions and recommendations. So how did the Agency fair in the discussion of these themes and have they been properly incorporated into the RTC?

Default Options - The RTC contains a good discussion on when it will consider the use of default assumptions in the screening phase and refinement phase of the residual risk analysis. For example, the discussion in the RTC about when EPA will consider the use of alternative approaches to the current cancer risk assessment methods which assume linearity at low dose levels to estimate cancer risk. The RTC provides an adequate discussion on using the principles outlined in the 1996 proposed revisions to the 1986 cancer guidelines. In addition, Congress is directed throughout the RTC to the recent and numerous Agency proposals which provide principles, uncertainty considerations and refinements to the many aspects which need to be considered when conducting a thorough risk assessment. Individuals who require specific examples of how the " nuts and bolts" of the overall RA and RM process are applied in any given situation will have to read referenced reports.

Data Needs - The RTC identifies the appropriate data needs in section 3.3. This section could describe the ongoing public and private research agenda, timetables and how the Agency will be assembling and evaluating information collected under other statues, such as the Toxics Substances Control Act (TSCA) to fill the data gaps associated with potential health and environmental effects of individual HAPs and HAP mixtures. The CRARM Report (pg. 126- 128) has a strong emphasis on the better use of the information collected under the Toxics Substance Control Act (TSCA) for making good risk assessment decisions. I have attached reports prepared by four State environmental agencies about the utility of information collected under TSCA which has been declared as confidential business information. The sharing of information between the Federal and State Governments and Industry is critical to the success of any residual risk program.

The RTC strongly emphasizes the lack of developed methods and ecotoxicity information for making adverse environmental effect determinations. Overall, the RTC adequately identifies numerous data gaps in acceptable risk assessment methodology which will make it difficult to depart from conservative default assumptions in some cases.

Validation (Methods and Models) - The RTC discusses the need for validation of the modeling assumptions used in the residual risk assessment program through the development of an improved model (e.g. TRIM) for use in the residual risk program, evaluations of existing state air toxics programs, and the ongoing data gathering effort to improve emission inventories and emission profiles from the source categories subject to

§112 (f). A discussion on the attributes of the TRIM model should be discussed in greater detail in the report.

Uncertainty - Covered very throughly by Dr. Frey.

Variability - There is a brief discussion in the report about concerns for sensitive subpopulations which could be expanded to account for individuals with preexisting diseases, multiple chemical sensitivity and other genetic factors which may lower the threshold for health effects for noncarcinogenic effects.

Aggregation - The RTC discusses additivity of risk and the multi-pathway evaluation of all other relevant routes of exposure. A very conservative approach , target organ and mechanism of action considerations may be needed in further iterations of the RA .

The RTC follows the overall recommendations of the NRC Committee by conducting conservative screening analyses in an iterative manner, and the introduction of refined methods and models in order to reduce the uncertainty in the screening risk assessments. It also highlights the opportunities for discussions with stakeholders throughout the residual risk decision- making process for the source category or specific facility.

The RTC follows CRARM recommendations for risk management across the board in most cases. It discussed the need for stakeholder involvement and participation in the RR determination process, the need for RA iteration and refinement process and guidance for making residual risk management determinations for emissions of known, probable or possible carcinogens. It also provides a framework for making residual risk management decisions for non-carcinogens through a hazard index approach, although the specific criteria for evaluating the public health significance of non-cancer effects have not be specified in the RTC. It properly recognizes the limited availability of guidance for assessing adverse ecological effects and the lack of consensus among the scientific community about what constitutes a significant ecological effect.

Charge Element 2. **Does the Report identify and appropriately describe the most relevant methods (and associated Agency documents) for assessing residual risk from stationary sources ?**

Dr. Medinsky response to this charge was thorough. My only comment as discussed at the meeting was the need for the assessment of acute effects induced by HAPs. The majority of the values in IRIS are for chronic exposure and the residual risk assessment with be made using these values. The Agency is going in the right direction concerning the need for acute values, but should develop these values on a selective basis. For example, a formaldehyde is a HAP which is in need of a chronic (cancer considerations) and an acute (upper respiratory irritation) reference concentration. As I discussed, there are times in

which various processes with emit high concentrations of HAPs over a very short period of time. These sources will generate complaints that are acute in nature (eye irritation, shortness of breath and in some instances possibly trigger asthmatic attacks), but will still be within the acceptable annual reference concentration due to the averaging of the emissions over 8760 hours.

Charge Element 5. **Does the Report adequately address the range of scientific and technical issues that underlie a residual risk assessment?**

The Report contains many of the health risk assessment protocol requirements that are required to be addressed by the regulated facilities in New York State. It contains many of the principles used in our existing conservative risk screening program (Air Guide-1) and provides a mechanism for more in-depth or refined application of risk assessment methodologies in an iterative manner. These types of iterative risk assessment have been done for specific source categories (e.g. MWCs) that have undergone a review under the State Environmental Quality Review Act (SEQR) for a determination of public health and environmental significance. The RTC contains a descriptive process for public involvement beyond public notice requirements in accordance with the recommendations of CRARM about stakeholder involvement and provides a good overview of the range of scientific and technical issues which underlie a residual risk assessment.

Overall the RTC emphasizes the dynamic and evolving nature of the risk assessment process and makes an attempt to limit constraints on the process by not being overly prescriptive while providing some bounds to the process in both the areas of RA and RM. This is an important feature of the process and the authors of the report should be commended for not creating a one-size fits all RA and RM cookbook. The process discussed in the RTC will allow for the continued evolution RA by allowing an avenue for the incorporation of recent advances in risk assessment science by endorsing the use of the iterative process.

A basic question for the SAB to decide is: **How conservative should the first risk assessment screening tier be?** The RTC provides a very conservative first tier screening assessment for public health protection. We currently use the MEI at the fence line for risk screening purposes in NYS. However, this MEI is an inhalation only MEI who is assumed to have an inhalation rate of 20 m^3 and weigh either 65 or 70 kg. In some cases this may be very conservative and in other instances it is not. For example, a review of the permitting decisions made for mercury emissions from municipal waste combustion facilities through a MEI site-specific multipathway exposure analysis did not result additional mercury controls with the exception of one facility. The multi-pathway health risk analysis for these facilities did not exceed inhalation or oral reference concentrations for mercury at the time they were permitted in the 1980's. Effects on wildlife were not considered, nor was the larger picture of the continued loading of mercury into the regional environment from the total number of these facilities located throughout the

northeast. The one facility which was required to put on additional mercury control was in an area which already has a serious mercury contamination problem due to past industrial activity. In this case, through the public process, the stakeholders (citizens) demanded additional controls and the final SEQR ruling required additional controls.

Summary notes : allows for Iterations (yes)/ Continues the use of public health and ecological conservatism in light of large uncertainty (yes) / provides an acknowledgment of inherent conservativeness of screen model and exposure assumptions (yes)/ provides for the influx of new RA methods and science as they become available (yes)/ provides a decision tree matrix to be used by risk managers (yes for carcinogens, yes for noncarcinogens as per CRARM report, not well defined for adverse ecological effect determinations).

APPENDIX A-5

Dr. Philip Hopke

The first charge is the determination of the correspondence of the approach to risk assessment with the recommendations of NRC Committee on Risk Assessment of Hazardous Air Pollutants and the report of the Commission of Risk Assessment and Risk Management.

A major problem in reviewing EPA's approach to residual risk assessment is that although the framework appears to be generally reasonable, the critical problems come in the implementation of the process in a real case and how the typically limited information is utilized and presented. Thus, it is hard to determine their adherence to the prior recommendations without seeing a worked example. The review of the 1994 NRC committee report does reflect the committee's major recommendations with respect to an iterative, tiered approach with uncertainty and variability. The summary indicates the need to document default assumptions and provide rationales for making specific choices. However, until the process has been applied, it is hard to determine the extent to which the recommendations will be followed.

Additional comments;

An earlier NRC committee that reviewed advances in assessing human exposure to hazardous air pollutants had suggested important changes in the approach to exposure assessment. The emphasis was to move to the examination of the distribution of exposures and away from unrealistic upper bound estimates for most exposed individuals. Although the document does indicate a willingness to eliminate the concept of the Most Exposed Individual (MEI), it still uses an upper bounding estimate, the Maximum Individual Risk (MIR), as the estimate of the person most highly exposed. In a context where costs and other considerations can be included, it is more reasonable to develop distributions of exposure and risk and then choose an appropriately high point in the distribution to perform the analysis on the basis of the likelihood that there will be a person who is actually at that risk. In general the bounding estimates still represent unrealistically high risks that no real individual is likely to actually incur. Thus, as a first tier estimate to eliminate the need for further analysis, the MIR would be acceptable, but better estimates are needed if regulatory action appears to be needed.

A major problem is the failure to validate models. The report indicates that there is still not validation of HEM and it is not clear to what extent new models like TRIM will be validated. It appears to be a common problem at EPA to develop models that are inadequately tested and validated before they are applied to regulatory decisions. There needs to be adequate testing and validation of any model before applying it to actual problem solving. It seems very unlikely that they can develop, test and validate a new model within the time frame available.

APPENDIX A-6

Science Advisory Board Review of Draft Residual Risk Report to Congress
Health aspects: Michele Medinsky

Charge element 2. Does the Report identify and appropriately describe the most relevant methods (and their associated Agency documents) for assessing residual risk from stationary sources? See especially Chapter 3, including discussions on health effects, dose-response, exposure, and ecological effects assessment. See also Chapter 4, screening and refined assessments (pp. 103-122).

The Agency has developed a well written, clear report that outlines a very ambitious strategy for assessing residual risks as mandated by the Clean Air Act. Assessment of residual risks for a broad spectrum of endpoints as a result of exposure to mixtures of chemicals arising from multiple pathways is a daunting task. Increasing the difficulty of this task are the following three issues: many of the methods proposed by the Agency to assess these risks are in the development stage even in the application to single chemicals; our toxicology knowledge of complex issues such as the potential additive or interactive effects of chemical mixtures at low doses and the modes or mechanisms of action of the individual HAPs is incomplete or rudimentary; and the data base for developing and validating models and assessing toxic effects is incomplete or absent for many HAPs. Communicating the limits of our knowledge and risk assessment tools to Congress in this Report is essential in order to prevent the misconception that we know more than we do. Congress and the public should not place an inappropriate level of confidence on the results of the residual risk analyses.

Because of the complexity and comprehensiveness of this risk assessment the Agency had elected to conduct the assessments in stages using a tiered iterative approach. Screening assessments will be used first. These assessments will likely use default assumptions and conservation models. If there is no significant residual risk, no further regulatory action is necessary. If however, a screening assessment indicates the risk may exceed a predetermined value, then more refined risk assessments will be conducted. This is an excellent approach to conserving limited human resources. Communication to all stakeholders regarding the conservative, screening nature of the assessment is critical, so as not to result in misinterpretation of the process.

The Agency presents a picture of the residual risk assessment process in broad brush strokes, as almost an idealized view of the process, with the underlying implicit assumptions that modeling strategies are in place, data needs are fulfilled, and knowledge of mechanism and modes of action are complete. However, the actual situation is much more complex, and many unknowns are subsumed into the details. In short, translation of the principles, as laid out in this report in to practice for the various individual risk assessments will be

fraught with unknowns. It is incumbent upon the Agency to present these unknowns in a thorough, straightforward manner.

The discussion on page 23 of the need for risk assessments for acute noncancer risks as part of the residual risk program is not clear. The Agency notes that "many HAPs also cause toxic effects after short-term exposures lasting from minutes to several hours. Indeed, for some pollutants acute exposures are of greater concern than chronic exposures." Intuitively, based on dose-response principles in toxicology, it would seem that a standard based on chronic exposures would protect against potential toxic effects due to acute exposures. However, this concern for acute effects likely rises because the acceptable exposure levels for HAPs will be averaged of a year. Thus, there could be periods of relatively high exposures followed by much lower exposures. If this is indeed the situation, then it should be discussed some in the report to put the acute exposures in context.

The draft acute methods document is an example of an important risk assessment methodology that is not yet in place. In particular, this document should harmonize, to the extent applicable with the EPA document for assessment of non-cancer effects due to chronic exposure (RfC methodology). For example, the dosimetric adjustments described in the documents are different at this point in time. Since both methodologies are assessing noncancer health effects, even though the toxic endpoints might be different, it is logical that both documents describe similar dosimetric adjustments.

A second issue regarding risk assessments of acute health effects relates to the usefulness of categorical regression in setting points of departure for acute effects. The discussion on page 29 is an excellent example of the theoretical nature of the residual risk report. This section presents a plan of action and an overview of the concepts underlying categorical regression. While there is little argument that if this methodology could be implemented it would be extremely useful in being able to simultaneously evaluate both concentration and duration, the methodology is not widely accepted and it is very likely that for many HAPs the data base is not sufficiently robust to implement this methodology.

In the discussion of chronic non cancer effects the Agency notes on page 27 the use of the Benchmark dose approach as an alternative to the NOAEL approach as a way to identify a dose without appreciable effect based on experimental data. The Agency's acceptance of the Benchmark dose methodology is viewed as a very positive step forward. However, these is still some question as to the Agency's application of uncertainty factors to the Benchmark dose. For example, most recently the Agency has applied an additional uncertainty factor based on the fact that the Benchmark dose is based on a finite response level, the theory being that this procedure is equivalent to converting a LOAEL to a NOAEL. However, this additional uncertainty factor is not universally accepted as being the appropriate approach. The appropriateness of the routine use of this uncertainty

factor is another example of where there is still some flux regarding the guidelines to be used in the assessment of residual risks.

In its discussion of cancer effects on page 30 the Agency notes that "If animal data are used in the dose-response assessment, a scaling factor based on the surface area of the test animals relative to humans is used to calculate a human equivalent dose. Surface area is used for this scaling because it is a good indicator of relative metabolic rate." However, differences in the rates at which humans and laboratory animals metabolize xenobiotic chemicals (including many of the HAPs) do not correlate with basal metabolic rate, and by extension the surface area scaling factor. Thus, surface area may not be a good indicator of the effective dose for chemicals that are metabolically activated. This factor should really be referred to as a default value used in the absence of specific chemical data.

Another area of uncertain methodology in the estimation of residual risks relates to assessing risks of mixtures. The current guidelines, first published in 1986 are currently under revision. Thus, it is not known how significantly the procedures for assessing risks of mixtures will change, although the Agency is to be commended for revisiting those guidelines. The Agency's proposal on page 61 to calculate a Hazard Index "for all components of a mixture that affect the same target organ using the RfC (even if the RfC was derived based on an effect in a different target organ)" is confusing and requires further explanation. As stated it appears that an RfC for based on a lung effect, for example, could be combined with an RfC based on another organ effect such as liver to obtain the Hazard Index. An example of how this index would be applied in a specific case would be useful. Additionally, on page 62 the Agency notes that "general additivity would include addition of effects that occur in different target tissues or by different mechanisms of action." The Report should make it very clear that the approach of combining chemicals with different mechanisms of action is purely a conservative calculation of maximum level of risk and not a process that is based on science. Ideally, additivity should be based on consideration of commonalty of mechanism; if chemicals do not act through a common mechanism their risks should be considered independently. This dependence on common mode of action for aggregating risks should apply to cancer and noncancer endpoints.

Charge Element 3. Does the report provide an adequate characterization of the data needs for the risk assessment methods? See especially Chapter 3 (pp. 50-63) and Chapter 4 (pp. 103-122).

In the Executive Summary the Agency notes that "Information available on actual health effects resulting from exposures to air toxics is limited." The Executive summary is an excellent place to introduce and expand upon the critical concept of a limited data base since many individuals may only read the executive summary. Additionally, references to uncertainty are found in other parts of the document such as on page 22 in the figure entitled "Sources of Information for Hazard Identification." However, the Agency should

be much more direct and thorough in explaining to Congress the extent of the data gaps and the consequences of the data gaps in terms of both the magnitude of the uncertainties associated with the risks and the level of confidence in the risk assessment. The quality, accuracy, and completeness of the risk assessments will depend upon the quality, accuracy and completeness of the data used in the risk assessment. The Agency should expand significantly on the issue of the data needs for conduct of the residual risk assessments and acknowledge the widespread data limitations. Limited data combined with default assumptions can result in risk assessments that are not well informed and that extend well beyond the boundaries of the underlying science. The impact of the data quality on the confidence associated with a guidance level has been addressed previously in the Agency's RfC guidelines where a descriptor is given for the confidence in the data base. The Agency could use that discussion as a starting point for text in this report that would inform Congress as to the limitations of the residual risk strategy in practice. A thorough treatment of the data base available for the conduct of the residual risk assessment would begin to inform Congress as to the complexity of the task at hand. This treatment should be highlighted in a separately identified section.

A good starting point for the development of a section on "Data Gaps" might be a table listing the current HAPs and some assessment as to the completeness of the toxicity data base for each of these chemicals. A good starting point for this table might be the table listing the HAPs in "Science and Judgment in Risk Assessment." Are there adequate chronic studies for assessing carcinogenicity, developmental and reproductive toxicity, and neurotoxicity? Are there any structure-activity indications that a chemical may have toxic effects that would not be manifest in conventional toxicity studies due to the lack of sensitivity towards these endpoints (e.g., immunotoxicity, respiratory tract hyperreactivity). Even for chemicals for which there is sufficient data for classification as a carcinogen; is there sufficient data to determine the mechanism or mode of action? Is there sufficient data to provide mode of action information for all the HAPs for all toxicity endpoints that could be used in aggregating risks for determining residual risks from mixtures? A table summarizing the data available to the Agency for assessing residual risk would enlighten Congress as to the difficulty of the task and the potentially large uncertainties associated with producing quantitative estimates. Additionally such a table would forewarn stakeholders at an early stage as to potential data gaps that could be addressed by either the conduct of new studies or bringing existing studies to the attention of the EPA.

Consistent with the need for as full a data base as possible for the development of residual risk assessments, the Agency should consider expanding its sources of useful data beyond that contained in its own data bases. High quality published information that may be critical for a risk assessment, or may be useful supporting information, may be so recent in nature that it is not in the EPA data base. There should be some mechanism by which new data could be brought to the attention of the Agency for inclusion. Likewise, early publication of significant data gaps in the development of the residual risk assessments could provide

an incentive for the rapid generation of the appropriate data by stakeholders or allow stakeholders to bring additional data to the attention of the Agency.

APPENDIX A-7

Comments on the Residual Risk Report to Congress
(April 14, 1998 Draft)

D. Warner North

Prepared for the Residual Risk Strategies Subcommittee
EPA Science Advisory Board
August 3, 1998

Revised: August 4, 1998

General Comments

My overall reaction to the draft report on Residual Risk (RR) is quite favorable. I find that the report presents an approach to risk assessment and risk management that is responsive to the requirements of the law (Section 112 f of the Clean Air Act, as amended in 1990). The draft report is also responsive to the recommendations in the 1994 National Research Council Report, *Science and Judgment in Risk Assessment* (hereafter, S&J: my assignment), and also, in my judgment, to the main thrusts of the reports of the Commission on Risk Assessment and Risk Management (CRARM). My comments below are intended to help in the process of refining and improving the current draft report. Some comments address the need to clarify language in specific sections. Others are motivated by a desire to see main themes in this draft or in S&J set forth at more length or with greater clarity.

Reflecting on the meeting, I believe there is a broad consensus among the SAB RR Subcommittee that EPA has EPA is to be commended for its effort in producing a framework that incorporates much of the guidance provided by the S&J and CRARM reports. Our criticisms address details of implementation and the need to go even further toward a flexible, iterative, and tiered system.

My main points of criticism, mostly related to S&J, are as follows:

More is needed on human health risks, and especially cancer risks. The discussion of ecological risks is overly long and detailed. I strongly support including ecological damage as an endpoint from hazardous air pollutants (HAPs), but there seems to be too much emphasis on this endpoint compared to human health. Some of the detail received criticism from the ecology experts at our meeting, and I found much of this criticism persuasive. I became concerned that EPA would put too

much effort and scarce resources into ecological risk assessment for a large list of HAPs and source categories. I recommend adding a simple and judgmental "Tier Zero" screen and a problem formulation effort involving stakeholders to select a small number of candidate HAPs/source categories that merit a more in-depth Tier 1 effort. This Tier 1 effort will identify candidates for a Tier 2 analysis (see page 119, second paragraph of 5.4.4.).

Suggestions for ecological candidates. Persistent organics and metallic chemicals that bioaccumulate in food chains are the obvious candidates. While most combustion products degrade, there are a few that persist, such as dioxin. The metals in the 17 HAP classes (Exhibit 15, p. 102 should be examined to see if ecological effects at ambient levels in soil and food chains might be significantly elevated compared to background, including areas where these metallic elements are present as naturally occurring ores or as wastes from mining and processing. Mercury compounds, lead compounds, and dioxins/furans are certainly deserving of Tier 2 analysis, but these classes pose significant non-cancer human health risks at low levels and have already been the subject of extensive risk assessment efforts. The Air Office of EPA should not be redoing analysis that others have already done in order to carry out its Section 112 f obligations.

EPA should not suggest to Congress that ecological risk warrants a large fraction of the resources without some Tier 0 and Tier 1 analysis to justify this allocation. EPA should attempt to identify chemicals for which more stringent regulation may be needed to avoid ecological damage than the level of regulation that is appropriate to protect human health. The list of such chemicals is likely to be a small fraction of the HAPs – certainly less than 20%, maybe less than 2%.

I recognize that much has been written about cancer risk and relatively little about risk assessment methodology for ecological damage from HAPs. Nonetheless, it is my strong impression that **much more should be said** in a report to Congress about how EPA proposes to implement the iterative, tiered approach to risk assessment **with respect to the complexities of cancer and noncancer human health risk**. Much of the detail on cancer risk is perhaps best addressed in EPA's cancer risk assessment guidelines, which have been issued in draft form and are to be finalized in the near future. Nonetheless, the audience for the RRS report in Congress needs a good tutorial on the issues. EPA has lots of good material from S&J, CRARM, and its guidelines. I expect that most of the important RR regulatory decisions will be on carcinogens that are judged to be linear at low doses, plus a few non-carcinogens like lead and mercury that can cause adverse human health effects at low levels of exposure. These substances will need very carefully done, high tier risk assessments as the basis for residual risk regulatory decisions. Congress needs to have an understanding of how EPA will do these risk assessments, including the level of effort needed and the importance of further data collection and research. An example

of an obvious omission in the draft report is lack of material on pharmacokinetics and biologically based modeling, which is barely mentioned in the RR draft.

Perhaps the most important need is to explain to Congress the large uncertainties and judgmental basis for cancer risk numbers in default assumptions such as low dose linearity, and the importance of these issues for risk assessment. See S&J, Executive Summary, first and third bullet at top of page 10, and Appendix B of the draft report, page B-3, first new paragraph. See also the extensive discussions in the two volume CRARM report. It is particularly important to acknowledge the uncertainty regarding whether the dose-response relationship for carcinogens (and some non-carcinogens) at low doses is linear or nonlinear. This uncertainty is going to be critical for many of the regulatory decisions on HAPs. The uncertainty and the underlying science should be clearly explained to decision makers and Congress, and not masked in discussion of complex risk assessment procedures such as benchmark dose and the linearized multistage model. The discussion should be transparent and readily accessible to the non-risk specialist. I am urging further efforts on a document that already represents significant progress from many preceding EPA documents on use of risk assessment in support of risk management decision making.

More is needed on the S&J recommendation that EPA improve its criteria for defaults and for departure for defaults. This issue is discussed at length in S&J and motivates some of its most important recommendations, in Chapters 6 and 12. While the issue is mentioned on page 10 of the draft, it is not developed adequately. A reader from Congress unfamiliar with cancer risk assessment might not even know what the National Research Council was talking about, since the concept of a default option is not introduced and explained.

More emphasis is needed on setting priorities for research and further data collection as an output from the iterative, tiered approach. The statutory need for residual risk assessments under Section 112 should provide motivation not only for EPA, but also for industry and other government agencies (e.g., NIEHS) to carry out needed research and data collection. Again, S&J is quoted on page 10 and Exhibit 1 reproduces the S&J figure that derives from the Red Book Figure 1, but the ideas are not developed.

Case studies are very useful to demonstrate how an iterative, tiered process actually works. EPA's benzene decision (Appendix B) is helpful in this regard. S&J provides several useful case studies in Chapter 6 and in its appendices F and G. These and other case studies (see, for example, Dennis Paustenbach's book of readings, various publications in *Risk Analysis*) should be cited. Case studies illustrate the issues in risk assessment, and how iterative, tiered risk assessment is carried out. The need for case studies was noted by most of the RR Subcommittee and the commenters from industry. It is reassuring that EPA plans to assemble such case

studies for a subsequent volume. For this draft report, more illustrative material from the published literature would greatly assist readers in understanding the framework that EPA plans to use for Section 112 residual risk assessment. It is not necessary to provide a lot of detail, but it would be very helpful to illustrate how the system will work using specific chemicals. The benzene NESHAPS as Appendix B is useful but much more is needed.

Specific Points in the Text

Page ES-3.

The text from the 1989 benzene NESHAP preamble is very important, as it forms the basis for the EPA residual risk strategy. The flexibility inherent in the words "approximately" and "ordinarily" needs to be emphasized, and the key areas of "science policy assumptions" (i.e., default options) and "uncertainties" need additional stress and explanation. Readers need to understand that cancer risk numbers (such as Congress wrote into the CAA Section 112 f in the 1990 amendments) are not precise, and that risk managers should have the flexibility to evaluate risks based on both quantitative and qualitative information. **Risk management decisions should not be strictly driven by the numbers.** Rather, the one in ten thousand benchmark should be an **approximate** guide to acceptability.

The severity of the endpoint (e.g., non-melanoma skin cancer that is readily treated and rarely fatal, vs. melanoma that is usually fatal if not surgically removed prior to metastasis) should modify what numerical level of risk is acceptable. This principle applies for both cancer and non-cancer health endpoints, and the policy stated in the RR draft report only addresses the cancer endpoint. Value judgments on risk acceptability are matters of policy, not matters of science. EPA's history on risk assessment for ingested arsenic illustrate the difficulties of incorporating such judgments into the risk assessment process. (See the Risk Assessment Forum document on arsenic, 1988, and subsequent SAB reviews.)

Page ES-4.

S&J, pp. ES-14,15 uses "iterative" and not "tiered" in its discussion, but the text of S&J makes clear that the recommendation is for both iterative and tiered risk assessment, as EPA is advocating in this draft.

Page ES-7.

The discussion of the IRIS data base should note the importance of achieving high quality in this data base through adequate budgets and internal and external peer review. See the discussion and recommendations in S&J, chapter 12 (pages 250-1, 265). The Air Office needs to work closely with the Office

of Research and Development to assure adequate quality in IRIS, and the expansion of IRIS to become a risk assessment data base, not just a toxicology data base.

Page ES-11, paragraph 2, line 5.

This text needs rewriting for consistency and clarity. I suggest: "… are integrated to portray the extent of the risk and characterize uncertainties in the risk." The task is not "… to determine a risk exists." Compare the wording of the benzene NESHAPS rule cited above: first (new) paragraph of page B-3.

Chapter 1, page 2

The importance of the word "flexible" should have even more emphasis. The issue is not that EPA has ten years to do the assessments of residual risk for the list of HAPs, but that EPA needs to adapt to the needs of the specific HAPS risk management decisions. "One-size" risk assessment will not "fit all" the differing HAPS regulatory decisions.

Page 10, following the bullets

I believe that the wording is inconsistent with what NRC intended in S&J. Delisting source categories and eliminating residual risk are not the appropriate choices of words.

Page 21, Section 3.1.1, first sentence:

Hazard identification does not give a yes or no answer to "determine whether the pollutants of concern are causally linked to the health effects in question." Rather, hazard identification provides a classification based on weight of evidence. Only for a small number of chemicals do we have sufficient evidence in humans that a pollutant is causally linked to a health effect such as cancer. Usually the evidence for causality falls far short of being sufficient, especially for humans. For many HAPs EPA relies primarily on animal studies. EPA uses the default option that observations of a health effect in rodent tests indicate the potential for that health effect to occur in humans.

So this sentence should be reworded. See, for example, S&J, chapter 2, page 26; chapter 4, pp. 57-60.

Page 22, box, second paragraph, last sentence:

This is a similar problem to the preceding comment. The "conservative public health policy which assumes that adverse effects seen in animal studies indicate potential effects in humans" is a default option. According to S&J such defaults should be noted and explained, and exceptions should be made where an adequate scientific basis exists. For these exceptions the results from

animal studies do not indicate the potential for human disease, usually because different biological mechanisms are involved in the different species.

Page 26, second (new) paragraph, third sentence through end of this paragraph.

This material is important and needs an expanded discussion, with some illustrative examples, as in Chapter 6 of S&J. The recent document prepared for the EPA Office of Water by a committee chaired by Dr. Julian Preston of CIIT is one of the few recent efforts I know of within EPA to grapple with the issue of departure from the default of low-dose linearity, based on current (incomplete) knowledge of biological mechanism. (Ref: Eastern Research Group, Inc., Report on the Expert Panel on Arsenic Carcinogenicity: Review and Workshop, National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, D.C., August 1997.)

Page 35.

As was brought out in our subcommittee discussion, there are serious questions about model validation. I advocate the use of simple, transparent fate and transport models for lower tier risk assessments, and use of more complex models requiring site-specific data only as needed and as justified by data availability for the upper tier assessments. The accuracy of emissions data is an important limit on the accuracy of the risk calculations. The time period of exposure is important, and the model should be matched to meet this need. For chronic health impacts, annual average exposure may be needed, but for some acute effects (e.g., bronchoconstriction from sulfur dioxide) peak exposure levels – averages over one hour or even less - may be needed.

Page 48, first new sentence of main text under box.

This point is important for health risk assessments as well as for ecological risk assessments.

Page 50, first paragraph of 3.3.1.

See previous comments on p. ES-7 regarding IRIS. The known and identifiable gaps and deficiencies in IRIS, HEAST, etc. with respect to the HAPs should motivate priorities for further collection and further toxicological research. EPA should not wait for the risk assessments to begin the process of establishing these priorities. That task should be starting now, and EPA should inform Congress about what resources it will require – from EPA and from other government agencies such as NIEHS. The priorities can then be refined as the risk assessment process proceeds at various tier levels. Recall that the process should be iterative, meaning that the important risk assessments will be revisited and revised in support of ongoing risk management.

Page 51, line 6.

This is the first mention I found of pharmacokinetics, and it is not very helpful to the non-risk specialist. This discussion should be expanded and aimed at the right audience.

Page 52, 53.

Same comment as above. This discussion needs substantial revision to provide a transparent and non-technical introduction to the use of iterative, tiered risk assessment. Limitations of models and default options, and provision for the use of more detailed models and departures from defaults, should be explained, motivated, and related to the resources needed to carry out the Section 112f risk assessment mandate that Congress has given EPA. See S&J, especially Chapter 12.

Page 56, line 6, then lines 6-8.

Add “and documented” after “studied.” Risk assessments need to document the source of data, models, and judgments used. Uncertainties also need to be disclosed and documented.

Page 56, last sentence extending onto page 57.

Excellent point, which should be expanded into a main theme of this RR report. See my comment on pp. 52-53.

Page 57, second new paragraph, line 5.

Has EPA changed its standard daily water intake assumption from 2 liters to 3?

Page 59, top bullet.

My impression is that most health risk assessors view SAR as unreliable. It is used primarily for determining needs for further research, as in the TSCA PMN process at EPA. I would be very concerned about using SAR as a basis for ecological risk estimates.

Page 60, bottom paragraph.

This approach seems to me quite preliminary and untried. I recommend dropping the paragraph unless OAQPS has significant experience indicating that this approach is proving useful. The type of analysis described in this paragraph could be a huge sink for analysis resources that yields little risk information useful for HAPs regulation.

Page 61, Section 3.4.

As discussed at the meeting, many of us believe mixtures should have a high priority for data collection and further research. This section seems inadequate and should be revised. Additivity makes sense in many situations as an appropriate default option, but in other situations (radon, particulates) known synergisms should be included. More attention should be paid to the philosophy of iterative and tiered risk assessment – start simple and refine the risk assessments for mixtures, based on the importance for regulatory decision making. By all means involve toxicologists in this process! Superfund is widely regarded as not having done a particularly good job of risk assessment based on good use of toxicology data and judgment of experienced toxicologists, and the Air Office should not blindly follow Superfund guidance documents. Toxic equivalency factors make sense as defaults for lower tiers, but a better approach based on collecting the tox data may be needed for higher tiers. See S&J, p. 103 and the SAB report referenced there.

Page 62, line 7.

The term, “complete cancer risk assessment” is incompatible with the iterative, tiered approach recommended in this document. Rewrite this passage!

Page 62, end of top paragraph.

Additivity may not be conservative. Use data and judgment obtained from experienced toxicologists, including external peer review!

Page 62, first new paragraph, especially last two sentences.

This material is good. It needs examples to motivate it and expanded discussion. Most of the time the nonlinear carcinogenic mode of action will not be well understood, and little will be known about how much the low dose risk deviates from linearity. Thresholds are not observable in the laboratory, and scientists are just beginning to understand the complex biological mechanisms involved. The limited knowledge and uncertainties need to be explained to users of risk assessments. Screening based on MOEs exceeding 1000 may be a good approximate guideline for acceptability, but avoid making it a bright line and remember it is a value judgment.

Page 63.

I agree with the comment at our meeting that the HI approach may be overly simplistic. EPA should involve experienced ecologists in risk assessment in the same way as toxicologists for health risk

assessment. Very simple criteria should be used only for screening out obviously low risks. If the risks are high, interactions among chemicals may motivate careful modeling based on expert judgment on the specific aspects of the chemicals and the ecosystem. The same point can hold for health impacts – expert judgment may be needed on pharmacokinetics and pharmacodynamic mechanisms in upper tier risk assessment, instead of continued reliance on default options.

Page 65, 4.1.1, first paragraph.

Some work was done by EPA in the 1980s on the public health significance of air toxics, but the results are only approximate bounding estimates. Check with the EPA Policy Office for the IEMP studies and other efforts to estimate mortality and morbidity at the national and regional level. Ask Dick Morgenstern and Dan Beardsley for references.

Page 66.

Important material. See my comments on page ES-7.

Page 68, Exhibit 14.

I am concerned that this exhibit implies there will only be two types of risk assessment. There need to be many types, motivated by the risk management need and the available information on the HAP. Defaults such as additivity may need to be relaxed, and these choices should be made based on expert judgment for the specific HAP.

Page 69, second paragraph.

Here and elsewhere, external peer review is needed.

Page 69, third paragraph.

Generally good. Consider replacing "is" by "may be" before "necessary" at the end of the paragraph. Available resources and other higher priorities may imply that the additional analysis is not done, and the lines are approximate. See previous comments on acceptability under ES-3.

Page 69, fourth paragraph.

The HI may be useful for screening, but avoid mechanical application and review the borderline cases with EPA and outside experts. Recall the need for quality in the IRIS data base – see previous comments.

Page 70.

Important material. See previous relevant comments about documenting basis for risk assessment, need for peer review, departure from defaults, and considering both possibilities when the evidence that a carcinogen is nonlinear is ambiguous.

Page 75, top.

Implementation of MACT on the point sources may cause area sources to dominate. Disclose this and other background such as high levels in indoor air or non-anthropogenic sources in the risk assessment. Relate to following discussion in 4.2.2.

Page 77, second paragraph, on voluntary and incentive based approaches.

I endorse Gil Omenn's comments at the meeting on use of risk assessments to obtain insight on the relative importance of risk reductions, rather than meeting thresholds for acceptability. Less risk is better, and the public wants help distinguishing big risks from little risks. Risk reductions might be accomplished through incentives and voluntary actions, such as the EPA programs described in this paragraph. In addition, the

information on emissions in the TRI may be an important motivation. The recently established Environmental Defense Fund website is an effort to publicize TRI data and motivate sources to reduce emissions via public pressure. California's labeling requirement under Proposition 65 is similarly an effort to require disclosure and use the resulting adverse publicity to motivate actions by manufacturers and users of chemicals to reduce risk. See the discussion in the CRARM reports.

Page 81, last full paragraph – also bottom of 83, top of 84.

More discussion of epidemiology is needed. Consider adding a reference to the Federal Focus "London Principles" report. Consider opportunities for epidemiological studies in highly polluted areas outside the US, where adverse health impacts (or biomarkers indicating increased potential for adverse effects) may be more clearly evident.

Page 82-84.

Much good material here, including some use of illustrative examples! More focus on the main implications from these examples for the non-technical reader might be helpful.

Page 87, end of first full paragraph.

There is a clear need to disclose information on background as part of the problem formulation, but EPA should avoid getting bogged down in too much detail in assessing background. Use the iterative, tiered philosophy and decide how much effort on background is appropriate.

Page 90, last sentence in continuing paragraph, top; page 91, following bullets; page 92, last two sentences of first paragraph; page 94, fifth sentence of second paragraph.

These excellent sentences should go into the Executive Summary and the introductory portions of the report. If they only appear in Chapter 4 after 90 pages they may be lost !

Chapter 5.

I thought this chapter was on the whole quite good. See relevant comments about issues from preceding chapters.

Page 105, second paragraph, line 6.

Typo: Should be Exhibit 16.

Page 105, third paragraph, line 7.

Consider describing the risk assessment as sufficient for risk management decision making rather than "complete."

APPENDIX A-8

Comments of Dr. G.S. Omenn

Here are my addenda to my extensive comments previously sent and pasted below:

1. Reference for public health context: Omenn GS. Putting environmental problems into public health context. Public Health Reports 1996;111:514-516.

2. Stress multiple scenarios and models, matched as much as possible to best available and attainable studies and data sets
   Note Benzene NESHAP options A,B,C,D--include the meaningful public health measure of cases per year (option B), with a threshold of some integer, like at least one case for local populations, perhaps a larger number on a national scale, for de minimus risk.

3. Focus massively on the experience to date with setting and implementing MACT standards:
   crucial for credibility of the Agency in responding to CAAA 1990;
   useful for estimating, projecting, and demonstrating emissions reductions, exposure reductions, and over time risk (endpoint) reductions in the range of greatest benefit and most reasonable cost
   illustrate multiple pathways and multiple endpoints analyses in risk range addressed by MACT stds
   try out uncertainty analyses
   Don't treat residual risk as a totally different program from MACT stds, only a backup if MACT stds are insufficiently effective.

4. Explain, explain, explain limitations of the methods and limitations of the data and models. Lower expectations and push back timelines.

5. Examine default assumptions to moderate the stringency of the screening risk assessment and make the transition from screening to refined risk assessment more dependent on detailed data and models, rather than hugely different simple assumptions (e.g., move from MEI and MIR to 90th percentile of real exposures, as recommended by Risk Commission, and use same exposure approach for both).

6. Push hard for stimulating new studies of clear relevance, including inhalation studies in animals and humans, biomarker studies linking rodents and humans, direct assays of representative mixtures.

7. Insert boxes with some available examples; list chemicals which are emerging under new carcinogen risk assessment guidelines as rodent carcinogens not relevant for humans, or suitable for nonlinear analysis, or assessed with MOE. Same for human exposures. Likewise, real examples for ecological risk assessment, from existing Agency documents. [I'm not asking for development of wholly new examples.]

8. Insert more detail about state and local air toxics programs, analytic experience, and similarities and differences with EPA intended program. In general, make much clearer how 188 pollutants x 174 sources will be managed and be respectful of limited resources for studies and analyses and decision-making.

Meanwhile, I urge you again to consider advising top EPA Staff to develop a policy memo soon to alert Asst Adm for Air and Administrator/Deputy Adm of the larger issues lurking here:

- level of detail appropriate for this audience/these audiences;
- statutory language to be revisited: ample vs. adequate margin of safety, in relation to relative priority for section 109 and section 112 pollutants; 10-5 vs 10-6 for flexible bright line for risk management;
- desirability of promising iterative process and communication;
- use of MACT stds process and implementation to illustrative feasibility of the various methods at higher emissions levels;
- clear warning about lack of sufficient relevant data in many respects. Feedback from the 12th floor would be more helpful earlier than later.

Best wishes.

GIL OMENN

PS I'm not repeating key points already made below.

>Date: Mon, 03 Aug 1998 04:26:46 -0400 >To: DON BARNES <BARNES.DON@epamail.epa.gov> >From: Gil Omenn <gomenn@umich.edu> >Subject: Re: Comments to date >Cc: medinsky@ciit.org, warner@dfi.com, hopkepk@draco.clarkson.edu, frey@eos.ncsu.edu, bucket@equinox.unr.edu, greg.r.biddinger@exxon.sprint.com, tjgentil@gw.dec.state.ny.us, zimmrmnr@is2.nyu.edu, TBurke@jhsph.edu >In-Reply-To: <s5c1a667.007@RT-MAIL2.RTPTOK.EPA.GOV>

> >TO: DON BARNES AND SUBCOMMITTEE MEMBERS.

> Please make sufficient copies for members and others at Monday's meeting. >FRM: Gil Omenn

> > >REVIEW OF RESIDUAL RISK REPORT TO CONGRESS FROM EPA (4/14/98 draft) > >COMMENTS FROM GILBERT S. OMENN, Member, Subcommittee on Residual Risk, EPA Science Advisory Board, meeting in RTP 3 August, 1998

> > >GENERAL COMMENTS

> >The Agency Staff have prepared a well-written, clear document faithful to section 112 of the Clean Air Amendments of 1990 and consistent with numerous relevant EPA guidance documents, the National Research Council reports on risk assessment (1983, 1994), and the reports from the Risk Commission (CRARM 1996, 1997a,b).

Many sections would be clearer if boxes could be added citing specific standards for specific chemicals or classes of chemicals that illustrate Agency discretion in applying general principles and moving beyond the numerous defaults, as recent policy documents have promised.

While the Agency indicates that no legislative changes are recommended, I believe it is timely to work with the Congress and the various constituencies to reconsider the peculiar and now-dated distinction between "adequate" margin for section 109 criteria air pollutants for which NAAQS are generated to protect the entire U.S. population and the "ample" margin for section 112 "hazardous" air pollutants to which much more limited portions of the population are actually exposed. In 1970 there was an overwhelming preoccupation with cancer risks, and a general desire to reduce risks to zero; there was little attention to other life-threatening, serious, salient adverse health effects. We know better now, yet we still have a long way to go in applying comparable analysis and risk management approaches to section 109 and section 112 pollutants. There is considerable text indicating new ways of analyzing both cancer and non-cancer effects and risks, but no examples are given and few are known to date.

The text mentions prominently EPA's development of multipathway analyses and what the Risk Commission multiple context analyses. The Residual Risk strategy should include comparisons not only with section 109 air pollutants, but also comparisons with risk-based decision-making for pesticide, water, Superfund, RCRA, and other Agency programs.

In Chapter 4, the general framework has two major flaws: continued use of the justifiably ridiculed MEI for the screening assessment and continued use of 10-6 upper bound as the individual risk level that generally meets ample margin of safety, rather than the

10-5 level chosen and recommended to EPA by the Risk Commission after extensive discussion in public hearings (see below). This matter may require amendment of section 112(f), depending upon interpretation of EPA's discretion. I believe the MOE analyses will show how remarkably conservative even 10-5 upper bound levels are, compared with other important health risks regulated by EPA.

Finally, on the research agenda for implementation of the Residual Risk strategy, it is unfortunate that so few of the 188 HAPs have inhalation studies available. Already more than 7 years have elapsed since the enactment of 1990 CAAA, with little additional investment in such studies, but lots of investment in uncertainty analyses and policy analyses for guessing about the potential effects in the absence of adequate data. The proposed residual risk program will go on until at least 2010, so we should not let another decade go by without investing in appropriate experimental and clinical studies, including studies that specifically examine the similarities and differences between rodents and humans and the appropriateness of numerous exposure, dose, and human/rodent equivalency factors.

SPECIFIC COMMENTS

Executive Summary

> Crosswalk and text very good.

> ES-2, para 4: insert in parentheses or footnote the 7 HAPs regulated. > ES-3, "ample margin of safety": see general comment above. Relate to other serious effects.

> ES-4, top and section 303: include focus on "risk reduction", not just contentious debates about very uncertain estimates of absolute levels of risk at certainly very low levels. As emphasized in the text, the Risk Commission also treated public health (and ecologic) context, total exposure analysis and attributable risks for specific adverse health effects, and proactive engagement of stakeholders for technical inputs as well as perceptions of risk and practical questions to be addressed.

> Q&A format useful.

> ES-6, HI, line 4: insert after "studies", "and species"

> DR: the old dichotomy between cancer and non-cancer DR analyses should be described as such, and a sentence should be added highlighting the Agency's work to find ways to look at cancer and non-cancer effects by similar methods whenever appropriate or potentially appropriate. [After all, we have no proof or even strong theoretical bases for the presumption that there are definable "thresholds" for noncancer effects, given the intraspecies variation and interactions with multiple other risk factors.]

> ES-7, IRIS: Need to discuss in text the widespread perception that many studies in IRIS were entered without adequate peer review and were retained even when new studies

showed different results. If this view is not justified, the text should take it on! Text does allow that external peer review applies to recent studies... [See also p.3-50]

> data: emphasize the paucity of studies by inhalation

> public health significance: might also mention risk management framework from Risk Commission, starting with putting each environmental problem into public health context.

> ES-8, para 1: awkward to say that screening method can "determine" whether continued emission of HAPs poses a risk; screening is not the same as determining...

> f,a,B: awkward that the Agency still will not, or cannot, cite any analyses of residual risk for even the earliest of the many MACTs issued over the past few years. [See also 1-2/3.]

> f,a,C: key gap in knowledge is inhalation route

> ES-9: sad that no epidemiologic or surveillance studies have been mounted or are proposed, yet policy still reads to rely on "available" data...How about proposing to join with public health agencies at federal, state, and local levels?

> f,a,C, background concentrations: disappointing to continue to rely on analyses of "incremental risk" of a particular source or activity, rather than estimating emissions/exposure/risk reduction and attributable contribution to reducing adverse health effect

> ES-10, negative consequences: have such analyses been done, or not, for MACTs? If so, name them in text.

> f,1,D: make clear that emissions are not synonymous with "problems"

> Chapter 5: goals should include reducing risks, not just estimating absolute levels of residual risks; example is radon from air versus radon from drinking water (forthcoming NRC report)

> "including all groups" is not as strong as "proactive engaging" groups/stakeholders, as urged in the Risk Commission report

> ES-11: The tiered approach is a big challenge for risk communication. It is hard to tell communities and environmental/consumer groups that the screening result indicates a potential significant remaining hazard and the rely on industry studies, generally, to conclude after "refined" analysis that there is no significant hazard after all.

> para 2: risk characterization includes not only toxicity and exposure, but also variation in susceptibility and exposure in identifiable population subgroups

1. Introduction

> 1-3: See above re: ample margin of safety; hard to justify in light of accumulating evidence of potentially lethal effects of section 109 pollutants' NAAQS. If to be used, need to define "adequate" and "ample" and explain and justify the difference.

>2. Background

> 2-7: "effective MACT standards will reduce a majority of the HAP emissions and much of the significant risk": (a) this statement is an analogy to the emergency removal phase of Superfund, which often greatly reduces estimated potential exposures and risks, yet is generally neglected in discussions of Superfund goals and successes or failures; (b) sure would be helpful at this point to have some quantitative data/projections for MACTs already issued and implemented.

> Descriptions of NRC x2 and Risk Commission are well done and sufficiently detailed to be useful to readers. The description of the Commission reports (2-11 to 2-16) appropriately emphasizes the risk management framework, the engagement of stakeholders, the early effort to put problems into public health and ecologic context, and the need to move from one chemical, one medium, one risk at a time to multi-source, multi-media, multi-chemical, and multi-risk analysis and management. Such contexts should be an explicit part of this residual risk strategy.

> The Risk Commission residual risk tiered approach is utilized in the proposed EPA strategy; the use of 10-5 as an action level was discussed extensively in public hearings and should be considered by EPA for its flowchart.

2-17: EPA guidelines..."as new information and methods become available" again sounds and is too passive.

> 2-19/20: Good to emphasize state and local air toxics regimens

3. Methods

3-22, box: Toxicologic data not "much easier to obtain", given huge deficiency of inhalation results. Ends with conservative old policy without mentioning EPA and Risk Commission efforts to identify mechanisms that are similar in rodents and humans, and those which are so different that risk assessment can be stopped at the hazard identification step.

3-24/25, boxes: again no mention of rodent/human similarities versus differences, unless implied under "limitations" in narrative statement. That's too obscure.

Amazing, in light of EPA's debates about ozone criteria document, that this document asserts that a "complete D-R relationship" can be characterized for ozone. Given the high background, it is quite uncertain what the effects of ozone might be in small proportions of the population (say, like 1 in 10,000 or 1 in 1M people) with much lower ozone concentrations...Same for SO2, particles, etc.

3-26, para 3: what examples can be cited (in text or box)?

Last sentence para 5: should the extrapolation methods be allowed to continue to differ so arbitrarily?

3-27, last para: "order of magnitude" should be changed to "factor of 10", since many people are quite confused by the phrase order of magnitude and use it for all kinds of factors...

3-28, para 1: be absolutely sure that significant figures do not claim greater precision than justified by the least precise of the input variables; last sentence is excellent, but credibility is uncertain unless examples can be given (see 10,000 factor on 3/27)

para 2, line 5: again investment in statistical descriptions/models, but not in actual experiments to get better data.

3-29, end of first para: please consider putting examples in a box

3-31, first para: for how many of the 188, or for what illustrative HAPs, are sufficient mechanism of action data available to justify or at least investigate, use of alternative models?

last sentence: Risk Commission worked hard on this matter of mixtures and additivity. In general, we considered additivity to be highly conservative; in many cases, related chemicals will be competing against each other for access to a common receptor or other target molecule. [Good statement on 3-62]

3-35/36: Important acknowledgments about lack of predictive capability of ASPEN and lack of validation for HEM. Much hope for TRIM; if to be in use by year 2000, it deserves much more description and assessment here; it also is consistent with multiple contexts of Risk Commission risk management framework.

3-37: pathways*be sure to indicate desirability of estimating numbers of individuals in the subsistence farmer or fisher categories in risk assessments, not just probabilistic risks per huge population denominator. [Likewise, box 3-40]

3-38/39: what duration of exposure is assumed for subsister fisherchild and for infant imbibing breast milk? Hopefully not 70 years!

3-39: Risk characterizations that use excessively precise risk estimates and uncertainty estimates should be discarded by the Agency or rejected by the stakeholders as manipulative, whether intentionally misleading or just sloppy.

3-41, box #5, "distributions": distinguish between distributions based on real data and distributions that are simply models and assumptions

3-42: consider benefit-cost analysis of the undertaking of detailed probabilistic uncertainty analyses of low-level absolute risk estimates versus relying on knowledgeable qualitative narrative

3-48: good point about consequences of "repeated use of upper-bound point estimates"; revisit assumptions before claiming that the screening approach yields an "unlikely, yet plausible"...

3-54: Excellent example (para 2) of use of PAMS and explicit listing of the HAPs measured. The monitoring system has come a long way since Lave

Omenn recommended rational siting and collecting of monitoring data ("Clearing the Air: Reforming the Clean Air Act", Brookings Institution, Washington DC, 1981). Should additional HAPs be recommended for this monitoring scheme? Indicate how many of the 17 categories are covered by these sentinel chemicals.

3-56, populations: mention also "genetic variation", especially in light of NIEHS Environmental Genome Project...by year 2010, there may be a lot more information, some of it relevant to residual risk estimation.

3-60: very unsatisfactory statement about mixtures. Surely, representative air samples with known multiple HAPs could be subjected to experimental studies to test whether the assumed additivity is even in the same ballpark with results.

3-60, Noncancer: Risk Commission also recommended use of Hazard Index in the Tiered Approach to residual risk and could be cited here.

3 -61, cancer/risk additivity: please note comments on mixtures and additivity above. Lots of criticism of the underlying concept was received by EPA in the comments on the dioxin documents cited here.

3-62: Good statement about mixtures; "possibility" of potentiation or synergism should be accompanied by "possibility" or "probability" of antagonisms.

"The MOE approach leaves the decision to the risk manager". That is exactly appropriate. There are so many uncertainties in the very low level risk estimates and there are non-scientific parameters that must be considered in risk management; this statement must be considered an important justification for trying the MOE approach!

Final para: Why is it not clear..? Surely one would be interested in knowing the LED10's and MOE's for serious effects of section 109 and section 112 pollutants. In general, data needs section (3-50 to 3-63) is thorough and good; needs to be tied to ORD program and other sources of funding for primary research, not just relying on "available data".

4. Other Statutory Requirements

4-65: How long can it continue to be true that EPA has no experience with assessments of public health significance or residual risk analyses? I thought several such analyses were well underway back in 1996-97. What about such analyses by academics or consulting firms, hopefully published??

** 4-66: The Risk Commission sought to give EPA a publicly-aired proposal for using 10-5 as the "bright line" for action after refined risk assessments, rather than the extremely conservative 10-6 (both upper bound risk estimates). EPA should reconsider the 10-6 action level proposed here. In fact, 4-67 says not 10-6 for each chemical, but 10-6 for the additive effects of the up to 188 HAPs; thus, some information from the MACT experience should be inserted to indicate numbers of HAPs per source category. Since these same facilities may be emitting section 109 pollutants, why analyze only the HAPs??

** 4-67, box: I am shocked that EPA would propose to continue to use MEI in its most extreme and most ridiculed form for the screening assessments. That is retrograde! Why reserve MIR to the "refined" assessment? This scheme is sure to confuse and ignite controversy.

5. Strategy

5-111 to 5-112:

Excellent to have explicit comparison with Risk Commission recommendations. However, the comments are quite brief. Hope public health agencies and academic scientists will be engaged in the process of putting air toxics problems into public health and environmental contexts. Could some reasoning be given for not utilizing the Commission's recommendation of 10-5 upper bound for the flexible bright line? [See comments at 4-66 and General Comments.]

APPENDIX A-9

Review of the "Draft Residual Risk Report to Congress"
August 04, 1998
George E. Taylor, Jr.
University of Nevada, Reno, NV
and
George Mason University, Fairfax, VA

This review encompasses previous comments offered in writing as well as comments offered at the subcommittee meeting on 03 August 1998. This reviewer supports the recommendation that the report generally meets the objectives as presented at the meeting by the EPA Staff. There are issues of clarification, emphasis, de-emphases, and technical accuracy that are important for the EPA Staff to consider in review process. These issues are outlined in this review.

Several general notations are in order. First, this reviewer endorses risk assessment and management as the appropriate methodology by which residual risk should be analyzed. This is appropriate for both human health and ecology. It is important to recognize that the methodology is not fully developed in all aspects, particularly with respect to ecology. Accordingly, the Agency is encouraged to further enlist the support of the SAB and other groups on an ongoing basis as the methodology continues to evolve.

Second, the document is clearly meant to be more of a guidance document, outlining in a general way the direction in which risk assessment will be approached for residual risks. Accordingly, there are a lot of unanswered questions regarding specific approaches to the analysis and management. Many of these issues must be addressed at some juncture, but it is not appropriate to do so here. However, my review raises some of these issues since they are critical and at least need to be on the table.

Finally, I encourage the Agency to recognize the role of ecology in its mission and to place a priority (co-equal to that of human health) on conducting the residual risk assessment for ecology. The tone of the report seems to apologize in an indirect way for doing ecological risk assessment in the first place and by so doing relegates ecology to a tertiary position. Some of those same concerns were raised at the meeting by the panel. There are many reasons for embracing ecological risk assessment as a valued party at the table, not to mention that the Agency has a legal mandate to do so and society continues to place a premium on ecology as a touchstone for quality of life. This issue is re-addressed later in my review.

1. Generality of Framework. The concern is simply the perceived sense that the report is general to the point of being "boilerplate"S. Too often, the discussion is couched in terms that are noncommittal, vague and elementary, and as a consequence it is difficult to ascertain what will be done. As it is now written, the strategy for the ecological risk assessment could take any number of trajectories, some of which might be "on target" while others would be well "Off the mark". To the extent possible, my recommendation is that these generalities be made clearer in the document. Many of these were discussed at the meeting and/or are presented herein.

2. Lack of Commitment to Quantitative Analysis. The document is replete with qualitative statements about the process and what will be done. In general, quantitative efforts are downplayed, and there is no significant discussion of which quantitative data will be gathered, analyzed and used in the risk assessment process. As in any effort, providing a boilerplate framework does not provide enough guidance on the quantitative aspects of the risk assessment strategy. It is important that the quantitative side of the risk assessment and risk management activities be formulated and dealt with in this document. Otherwise, the qualitative risk assessment may be viewed as what is expected.

My recommendation is to make sure that the quantitative rigor of risk assessment and management is given more visibility. The issue is not one of presenting the quantitative aspect in this report but making sure that the audience appreciates the degree to which quantitative analyses will be performed.

3. Ecological Underpinnings. The ecology sections of the document are oftentimes couched in very simplistic terminology and this may be problematic. To ecologists, there are sections that simply do not reflect accurately current ecological theory and practice. My recommendation is to re-visit these sections and have an outside ecologist offer extensive revisions with references. Many of the critical areas of concern are noted elsewhere in this review.

4. Second Tier Risk Assessment for Ecology. The second tier ecological risk assessment calls for a far greater and in depth analysis of risk. However, there is no road map that details the quantitative nature of this effort. Statements are offered that the second tier is more quantitative and accurate, but how that is done is glossed over in the report. My recommendation is that this report provides more guidance in what the second tier analysis will look like in general terms (including the degree of quantitative rigor and uncertainty). At a later date, it is important that an example of this second tier risk assessment be presented for review.

5. Economics of Ecological Risk Assessment. Any second tier risk assessment dictates that managers account for issues other than ones that are the domain of ecology per se. For example, managers must account for social and economic concerns. There is

no discussion of to whom these issues will be addressed either in a qualitative or quantitative sense; there is no discussion of the methodology to be used. My recommendations is that more articulation of the second tier risk assessment be offered in this report.

6. References. The report provides some general documentation as to what will be done and the methodologies to be used. However, the reference list is insufficient for my needs. I would like to see far greater referencing, particularly to the peer-reviewed literature. That provides some assurance that the strategy will be tied to prevailing and generally accepted principles in the scientific community. In this light, it is critical that the references be ones that are current.

7. Models. The role of conceptual models is overplayed in the analysis. While these may be of value, their role is largely qualitative as described in the report. They have value in risk communication but their value in risk assessment per se is limited. The use of quantitative/simulation models to investigate the behavior of ecological systems is not presented as an option in the analysis plan. I recommend that modeling as it has evolved in ecology play are far more prominent role in analysis. This approach is given wide use in the transport, transformation and fate sections. The ecological sciences has come very far in the development and use of simulation models to address effects on ecological systems, and the application of these methodologies is very appropriate to residual risk assessment. They are particularly appropriate for analyses that are at watershed and regional levels.

8. Mechanisms of Action and Mixtures. The analysis calls for the HI methodology for additive components. In ecology, we are a long way from knowing mechanisms of action for most pollutants. Accordingly, the default methodology will be what? It is unrealistic to default to a molecular mechanism of toxicity as the means of addressing mixtures (or individual chemicals) for ecological risk assessment. For most pollutants at chronic levels in ecosystems, the effect is largely mediated through some ecologically/physiologically significant process that governs fitness (e.g., photosynthesis in plants, respiration in animals, reproduction). My recommendation is that mixture (as well as single chemical effects) analysis in ecological risk assessment be based not on molecular mechanisms of toxicity but instead on how the pollutant affects critical processes governing fitness.

9. Background Concentrations. The argument is presented that background concentrations will be the additive combination of the following: (1) natural background concentrations and (2) any additional concentrations due to other anthropogenic processes. I am uncertain about the inclusion of the latter and would like to see more discussion form a pragmatic basis as well as an ecological perspective. This approach differs from that used for the criteria pollutants in which natural background is defined as that solely in the absence of human technology (to the extent

it can be determined). My recommendation is that the Agency re-address this issue and provide a rationale for the decision. The problem with the proposed method is that aggregation of chemical effects will be eliminated as a residual risk (i.e., effects of individual categories will be done in isolation from that of other concurrent sources which individually will not reach a threshold but collectively may exceed a threshold). Does this approach meet the intent of the residual risk analysis?

10. Decision Strategy for Risk Managers. The entire process of risk assessment is oriented toward managing the risk via risk management. The discussion is abbreviated on how risk mangers will make their decisions. There is no discussion of the approach that will be used for the manager to decide what to do and how to proceed. This is a critical step and needs to be articulated.

My recommendation is that the report provides a full section on the methodology that will be followed by risk management in the same general manner in which the risk assessment approach is articulated.

11. Stakeholders. Having stakeholders involved in the risk assessment and management process is appropriate. However, there needs to be some language as to how stakeholders will be identified and represented. This issue is important to articulate guidelines for from the beginning. My recommendation is that a general position be developed on the rationale for stakeholders, how they will be incorporated into the process (e.g., as currently stated, they will be solely in the risk management process; will their role be solely advisory?), and generalities of the selection process.

12. Comparative Risk Assessment. This term is used in a number of places in the document, and there is no definition of what comparative risk assessment is relative to other forms of risk assessment. My recommendation is to define the term.

13. a priori Screening of HAPs in Ecology. The document offers two criteria for identifying HAPS as potentially hazardous. The two are (1) potential for bioaccumulation and (2) lifetime. These are important, but I would argue that these two criteria alone are insufficient. Two examples illustrate the point. Ozone is one of the most significant regional pollutants affecting ecological resources and human health. The residence time of ozone in the atmosphere is minutes to hours, and its bioaccumulation potential is zero. Thus, ozone would not be identified in the initial screening exercise for ecology. In another case, CFCs have a long residence time, but their residence time is in the atmosphere (stratosphere) rather than the earth's crust or the biosphere. Moreover, the mechanism of action of CFCS is via UV-B enhancement, and it is not clear how this would be handled in the initial screen? In a related aspect, since many of the HAPS are a concern because of their transformation into derivatives that are also toxic, it is important that the derivatives are identified in the process (and this is done). If the derivatives are criteria pollutants (e.g., ozone,

nitrogen oxides), will they be handled as a residual risk? My recommendation is that the Agency develop a more robust approach to the first tier of screening criteria for ecology and ensure that the criteria minimize the probability of a significant false negative. The two proposed criteria are a start but need to be amplified. Other criteria might include inherent toxicity. Contribution to criteria pollutant levels, partitioning in the environment (e.g., KOW, deposition to canopies), etc.

14. Uniqueness of Ecological Risk Assessment. Risk assessment is presented as a generic methodology common to both human health and ecology. Because ecological systems are far different than human systems in terms of risk assessment and risk management, it is important that the unique aspects of ecological risk assessment be discussed explicitly so that managers and assessors know where the distinctions lie and how those distinctions need to be addressed differently. My recommendation is that a full section be devoted to this issue.

15. Hierarchy Theory in Ecology. The report clearly places an emphasis on ecological risk assessment at broad spatial scales (e.g., watersheds, regions, etc.). While there are a number of ways to investigate broad spatial scales, the report places an emphasis on the "Scaling Pup approach" in which data collected at the molecular and individual level are translated to higher scales (population, community, ecosystem). The report assumes without a discussion that analyses at one scale of hierarchy dictate applicability to scales higher up (or lower down). For example, is it consistent with hierarchy theory to state that protecting the most sensitive cohort dictates that the ecosystem will be afforded protection? I recommend that any discussion of trans-scale applicability of analyses be tied to where the scientific literature endorses that perspective. I am not convinced that those issues are resolved to the extent assumed in the report. It is noteworthy that one of the dominant shortcomings of the analysis of global climate change is exactly this problem (limitations of scaling up).

16. Analysis of Broader Scales in Ecology. The only technique that is presented to address broader scales in ecology (watershed, ecosystem, etc.) is the scaling up approach; that is appropriate and is one method. However, there are other methods that can be stand alone or used in conjunction with others. Examples include modeling, geographic information systems, ecological epidemiology, remote sensing, etc. My recommendation is that the report recognizes the multiple methodologies that might be used and that the scaling up approach is one of several supplementary and complementary tools.

17. Sensitivity of Plant Systems. Statements are made that plant systems are far more resilient and resistant to stress that animal systems. I am not convinced that this statement is accurate and reflects prevailing scientific knowledge. I recommend that this section be re-visited. If the report is convinced that this statement is accurate, perhaps a reference would be in order.

18. Death of Individuals. The argument is presented that society is not concerned about mortality of individuals. That statement needs to be re-visited. There are stakeholders that would very much disagree with that phraseology. Notable examples of recent studies include the panther in Florida, rare and endangered populations, species in highly valued ecosystems, etc. My recommendation is that the report recognizes that there are situations where mortality of individuals is important, particularly to some local stakeholders.

19. Information Sources for Ecological Analyses. A number of sources are identified for obtaining data and methods for risk assessment. Several are prominently missing. The first is the use of quantitative (not conceptual) models (see above discussion item). The second is the field of ecological economics. High priority is given to the role that economics will play in ecological risk assessment/management and yet there is no discussion of what economics will be used. The third is ecological epidemiology (see Item 16 above), I recommend that the report expand the list of information sources.

20. Ecological Significance Discussion. On page 118, the report outlines what defines ecological significance. The discussion needs to be re-visited to ensure accuracy, prevailing ecological theory and practice and society's view of quality of life. This section could benefit from a strong linkage to peer-review literature. One of the major omissions is the argument about the sustainability of ecological systems and the linkage between human well being and the functioning of ecosystems.

21. Concern Regarding Readiness. The issue of residual risk is appropriate for the risk assessment methodology. I endorse the Agency's position that even though not all of the methodology of risk assessment is fully developed in either the human health or ecology arena, the analysis needs to proceed rather than waiting until all the "i's" are dotted.

22. Use of the Maximum Exposed Individual. I endorse the use of the MEI as a screening tool in conjunction with other approaches. For the second tier risk assessment analysis, the role of MEI should be downplayed, with greater emphasis placed on more realistic exposures.

23. Scales and Residence Times in Ecology. The principle that residence time and distribution are critical aspects of the ecological risk assessment forces the issue of the spatial scale to be assessed. If the chemical has a residence time of a month or more, then the distribution of the chemical will approach hemispheric proportions. Longer residence times in the atmosphere will dictate global distributions. Will the scale of the ecological and human health risk assessment be scaled according to the atmospheric residence time? It is hard to justify for reasons of national interest but from a first principles perspective it is difficult to ignore the issue. I am not sure of the legal

mandate for this issue, but the Agency needs to recognize that the spatial scale for distribution may be quite large. The best example is that of mercury. Given that the residence time of mercury in the atmosphere is one year, and change in emission will influence mercury accumulation in the US to a very small extent. Will the risk assessment assume that a 50% reduction in emissions will translate into a 50% reduction in US accumulations?

24. Role of Ecology in the Agency's Mission. I am a strong proponent that the role of ecology should not be relegated to an addendum to human health issues.

First, the Agency has a mandate for ecology and that mandate is a unique feature of the Agency's charge relative to other organizations. Second, the intrinsic value of ecological systems is becoming (and will continue to become) a significant priority for communities and quality of life issues. Third, there is intrinsic and monetary value in having ecological systems functioning well (see recent literature on ecological economics). Finally, the well being of human systems is inexorably linked to the sustainability of ecological systems (issue of sustainability), so it is ill advised to disassociate one from the other.

My recommendation is that ecology not be relegated to a step-child initiative in risk assessment but rather that it be embraced as a co-equal.

APPENDIX A-10

Rae Zimmerman

Science Advisory Report, Residual Risk Subcommittee SAB RRS - Additional Preliminary comments on the Draft Residual Risk report for Charge Element 5: Does the Report deal with the full range of scientific and technical issues that underlie a residual risk program? From Rae Zimmerman 8/5/98

These comments combine and extend the two previous sets of comments submitted for the 8/3 panel. They address the depth of coverage within each of the topic areas, implied by Charge Element 5. References are made to a few of the comments provided by other SAB RRS members.

SUMMARY HIGHLIGHTS

-Provide a clear risk assessment/risk management framework for the residual risk program, choosing from among or synthesizing those presented in the report and available elsewhere

-Be clear about the charge, i.e., that it builds upon MACT, and therefore the existing or expected results of MACT should be set forth

-Supplement the EPA literature with a wider literature, including a process for referencing and explaining the prevailing opinion with respect to positions chosen for the program

-Since public health is a key focus of the residual risk requirements, it should be portrayed in flow chart form based on the material from p. 65 on.

-Specify procedures within the iterative screening process to a greater extent, incorporating, for example, a number of shortcuts like the use of standards or indicators that encompass others to avoid redundancy.

-A clear approach to and strategy for uncertainties should be set forth prior to embarking upon the details of uncertainty in specific contexts. Uncertainties that are likely to stop the whole process of residual risk estimation should be particularly identified. Uncertainty should not be a reason for dropping from consideration residual risks from a particular source or chemical.

DETAILED COMMENTS

INTRODUCTION

This Charge Element is distinguished from the others in that it is an accounting of how well the report covers or incorporates the various issues. In contrast, the other charges address the details of a particular issue. In that context, the range of scientific and technical issues encompasses:

· the depth and degree of detail of approaches and methodologies brought to bear on residual risk determinations for the components identified in Section 112 (f) (1) and · the breadth of the issues covered under Section 112 (f) (1), that is, whether all of the issues necessary to conduct feasible residual risk determinations are listed in the section, and what process was used in determining the elements listed in the section.

CLARITY OF OVERALL PURPOSE

The report is very comprehensive in its scope and coverage of current EPA methods for and approaches to risk assessment. The purpose of the report needs to be made clear, however - is it a review of current risk assessment methods or does it specifically extract from the literature what is applicable to a residual risk program? The statutory language is relatively clear in stating that the report is to "investigate and report" (and recommend) risk calculation methods, public health significance of estimated remaining risk, and actual health effects. The strategy is presented at the end of the report, and it should incorporate more of the elements mentioned throughout the text as relevant to the residual risk program. Whether or not and how residual risk assessments differ from other risk assessments should be identified in the introduction.

The report should be more explicit about what it is providing. For example, on p. ES-5 it states that Chapters 3-4 address statutory requirements and information on the methods. The report can't solve all of the outstanding risk assessment problems, and must be selective, focusing on what is relevant to residual risk. Instead of providing such a focus, these chapters are primarily reviews and critiques of what exist. Although they suggest methods, the suggestions are either not fully explained or are themselves critiqued back and forth so the reader often doesn't know what decision is being made for the residual risk program. The one exception is the public health analysis, which is very complex and would benefit from a flow chart. The report indicates that its focus is on programs, methods and approaches relevant to residual risk, but the distinction is not always apparent in the write-up, since it covers aspects of risk assessment applicable to everything.

PROBLEM DEFINITION: DIFFICULTY OF CONDUCTING RISK DETERMINATIONS ON MACT RESULTS

The report should clearly acknowledge that its charge begins where MACT left off. In order to build upon MACT, the report should summarize what has been accomplished under MACT. The strategy chapter very correctly points out that MACT control strategies and standards did not have risk in mind when they were developed. Therefore, in addition to building upon MACT, the residual risk program should identify as a first step in the residual risk program or strategy translating what was done in MACT into risk terms for comparability. The current approach to MACT is source-based (174 sources), and sources are generally defined as industrial groupings. The specification of 188 HAPs ranges from individual chemicals to chemical groups. Risk assessments are best performed on a specific hazard. Both the generality of the specification of sources and 17 categories of HAPs make a traditional risk assessment impossible, and when HAPs and sources are combined, the degree of generality magnifies.

This can be approached by increasing the number of scenarios, assumptions and correction factors, but these should be spelled out or at least identified as an initial task of the residual risk program or strategy.

Background Concentrations and Conditions. The role of background is a problem that surfaces throughout the Clean Air Act and other environmental legislation. An approach to the problem should be clearly a part of the residual risk program strategy. Background is most easily approached by defining particular location, time period, or reference source. The report should draw upon the experiences of other programs, and in particular, consider how background and baseline frameworks relate to one another.

RISK ASSESSMENT/RISK MANAGEMENT PARADIGMS

A number of risk assessment/risk management paradigms exist. Since these set the stage for the document, the models should be synthesized or a single one should be adopted. The report identifies two of these. The NRC "Science and Judgment" report (Exhibit 1, p. 11) provides one model and the CRARM provides another model (Exhibit 2, p. 13). These have to some extent been superseded by EPA's more integrative model adopted in the U.S. EPA ORD strategic plan (U.S. EPA, ORD, "Strategic Plan for the Office of Research and Development," Washington, DC: U.S. EPA, May 1996. EPA/600/R-96/059. P. 3. Also contained in the April 1997 update.). The EPA paradigm has arrows going in many more directions than the other two models implying a greater degree of integration of the various steps and stakeholders. This is important, because the report needs to come to terms with how it integrated stakeholders, i.e., in providing data, in decision-making, etc. Finally, the National

Research Council, "Understanding Risk" (1996, p. 28) report implies still a fourth model that is far more interactive and involves stakeholders to a greater degree than any of the others.

BREADTH OF THE KNOWLEDGE BASE FOR THE RESIDUAL RISK REPORT

The report relies heavily on existing regulations, guidelines and special commission and National Academy studies as a basis for its approach. An assessment of other knowledge is needed, and to what extent the existing documents mentioned in the report bounded or constrained the approaches taken toward residual risk. Over two-thirds of the references listed are EPA or other CAA required or commissioned studies. Although peer reviewed literature is, of course, contained in these documents, these documents did not only address residual risk, and the residual risk report could include literature specific to the residual risk concerns. Other RRS members also indicated that the references were limited and the peer-reviewed literature should be used. Even though the references cited incorporate peer-reviewed literature, more direct references are needed.

The use of a broader literature is particularly critical given several questions that have been raised about the lack of acceptance in the scientific community of some of the approaches advocated in the report. Medinsky, for example, points out that the use of categorical regression for acute effects and the use of surface area to extrapolate animal findings to humans may not be generally accepted. Taylor, furthermore, points out that the basis for ecological significance (p. 118) used in the report is not widely accepted. Leaving issues such as these unresolved questions the credibility of the report.

The report contains a number of factual statements that, although reasonable, need to be supported. For example, carcinogenic default assumptions include MOE (for non-linearity findings) and linear low-dose extrapolation where no d/r data exists. For non-cancer endpoints (chronic) - inhalation RfC/D is used as a scientific base. The prevailing opinion in support of these statements needs to be referenced or supported.

Thus, a means of not only drawing upon a wider literature, but also developing a process is needed to ascertain the prevailing opinions of the scientific community (if not consensus) on many of the issues brought up in the report.

A NEED FOR SIMPLIFICATION WITHOUT COMPROMISING SCIENTIFIC VALIDITY AND COMPREHENSIVENESS

The report authors should be commended in underscoring the use of an iterative screening technique as an initial step for residual risk. This approach is a useful means of simplifying the process of making residual risk determinations, and seems to have wide

support. The iterative screening technique would be strengthened by greater specification of the procedure, what it depends on, and how it is applied specifically to residual risk. There are a number of simplifications described below that could be applied to the residual risk program.

Nested sets. There are apparent short-cuts in developing a strategy for standards in general that apply to residual risk determinations. First, ecological standards in some instances can be more stringent than health standards especially where food chain effects are a consideration (unless the ecological standards are constrained by economic and technological considerations). Thus, where an ecological standard is called for, it can serve the dual purpose of protecting public health and the environment. Second, as Medinsky points out, chronic exposure standards inevitably protect against acute exposures for a given chemical. This implies that guidelines or standards for acute exposures can be precluded by chronic exposure standards, only of course where the time period specified for chronic effects encompasses time periods for acute exposures. Third, understanding the relationship between chemicals and their precursors can simplify the identification of potential risks. Related to this is the fact that chemicals are often found in association with other chemicals. Agent Orange, for example, was associated with a contaminant dioxin. Although Agent Orange was immediately suspected as being toxic, zeroing in on Agent Orange made it easier to identify dioxin as the real source of concern. Fourth, one approach to dealing with data uncertainties as well as simplifying the development of standards is to go through a progression of linked data sets. For example, the most specific data for an air toxic risk assessment would be knowledge of a known health effect associated with a known exposure. If that information is unavailable, one works back to exposure indicators. If exposure information is unavailable, one then draws on source-based measures, etc. These are conceptually linked data sets. (See, for example, R. Zimmerman, "Governmental Management of Chemical Risk," Chelsea, MI: Lewis Publishers (and CRC Press), 1990. Pp. 70-71.) The approach to residual risk needs to take advantage of these economies in the screening technique.

UNCERTAINTY

The report identifies numerous uncertainties and methods for dealing with them throughout the report. The report would benefit from an overall outline of uncertainties that distinguishes among the different types. These types include uncertainty generated from the absence of data, sensitivity of results to fluctuations in the parameters selected, values of parameters, and the structure of the equations that relate sources, exposure and risk. The one that is typically given the least attention is uncertainty arising from the wrong or insufficient choice of parameters.

"Fatal Flaws". The report identifies problems that may very seriously constrain the process of developing a residual risk program. For example, the report identifies the

serious limitations in the availability of actual monitoring data for air quality. "Presently, there is no national ambient air quality monitoring network making routine measurements of air toxic levels." (p. 53) The implications of this for both model development and validation should be addressed directly. Similar uncertainties arising from emissions data availability are identified as well, although this is at least handled through reliance on emission factors estimates.

Approximations. Recognizing data deficiencies and uncertainties, the report advocates a number of approaches, most of which are widely used. The first is default assumptions. This is not new, and is underscored in the "Science and Judgment" report but only if followed by iterations, continually revisiting results when new information is available. Some of the simplifications used in the report include:

Use of categorical regression results for acute effects (p. 29) Benchmark doses as alternatives to NOAEL Additivity for chemical mixtures Uncertainty factors Extrapolation techniques to convert animal test results to humans

It should be pointed out that approximation methods are better than dropping out a chemical or source from a risk determination just because there is no data on it. This problem will be more important for residual risk than it was for MACT. Statements like "Assessment endpoints that cannot be linked with measurable attributes should not be selected" (p. 45) could imply leaving out things that could be important but for which data is currently not available. One should remember that the EPA decision to drop a number of chemicals from being considered for safe drinking water standards in its regulations because of the lack of data became subject to judicial scrutiny.

Model Validation: degree of acceptance, testing and validation of models and methods. The report acknowledges the fact that a number of models have not been validated. It also does not seem likely that such validation will occur prior to the residual risk determinations. How will the residual risk program handle that?

THE EXPRESSION OF THE RISK CALCULATION

An area of debate identified in the report seems to be whether risk calculations should be expressed as point estimates or probabilities and ranges. In the context of ecological conceptual models, the report concludes that "the point estimate approach is most useful as a screening approach" for an unlikely, worst case scenario (p. 48). The report points out that probabilistic approaches while displaying a distribution of results, often cannot find distributions for input data and the results are hard to communicate. The use of point vs. probabilistic values is a fundamental philosophical issue that should be dealt with in a more general discussion of the residual risk approach.

DECISION PROCESSES

The role of stakeholders in the residual risk determination process is unclear in the various models presented. Potential roles are as a minimum in providing data, and more significantly, in where they come into decision-making. The models in both "Understanding Risk" and in the EPA ORD new paradigm for risk assessment and risk management underscore the thorough integration of stakeholders into decision-making.

PUBLIC HEALTH

Public health is a fundamental part of the residual risk program, specifically identified in the legislation, and it would benefit from a clear flow chart based on the material from p. 65 on. The report (p. 69) identifies specific numerical criteria for public health significance. Separate criteria are used for screening level vs. refined approaches, and thresholds differ for cancer and non-cancer endpoints. For carcinogens, the trigger for refined analyses is one in a million risk. Non-cancer risk will use hazard quotients, and a hazard index value of 1 would be a trigger for the refined analysis. The approach and the values should be justified.

UNINTENDED CONSEQUENCES

The legislative mandate for the report clearly requires the identification of negative adverse consequences resulting from the imposition of residual risk requirements. A framework that includes socioeconomic consequences as well as environmental ones should be set forth, drawing upon the methodologies used in environmental impact assessment and risk-risk comparisons.

MISCELLANEOUS ISSUES

Extreme values. The report occasionally lists the fact that "high-end" values rather than just average values be considered in the estimates of residual risk. This is notable and should be treated consistently throughout the various elements of the strategy.

Population estimates. What methods are used to identify nearby populations? Results can differ dramatically where approximation methods are used that are not health-based because no actual health data are available. Also, the issue as to which populations should be examined arises continually - workers, transients, etc.

Availability of epidemiological or other health studies. In searching epidemiological and health literature, an important consideration is whether or not the report should set forth criteria for matching literature conditions vs. those conditions relevant to a particular relative risk determination. Also, when such literature should be invoked is an important consideration, that is, where in the screening process it is relevant.

APPENDIX B

A MORE DETAILED DESCRIPTION OF THE SAB PROCESS

The SAB Staff recruited Dr. Philip Hopke, Dean of the Graduate School at Clarkson University, to serve as Chair of the Subcommittee. Working with the Chair, other SAB Members and Consultants, and Agency Staff, the SAB Staff compiled a list of over 40 scientists and engineers who were subsequently surveyed for their interest in and availability for participating in the review. The Chair and SAB Staff made the final selections for membership on the Subcommittee and assigned different members lead responsibilities for each of the Charge Elements. When informed at their July 18-19 meeting of plans to conduct the review, the SAB Executive Committee raised no objection to proceeding with the meeting.

Therefore, on August 3, 1998 the Subcommittee convened in the Main Auditorium of Environmental Research Center at the USEPA laboratory in Research Triangle Park, NC. Minutes of the meeting are available. Each member of the Subcommittee submitted written comments on the Charge Elements for which he/she had lead responsibility. Three members of the public provided comments on the technical issues under discussion. Following a full day of discussion Subcommittee members were given the opportunity to enhance/modify their written comments. These written materials were the basis for A Subcommittee report that was drafted by the Chair and the SAB Staff and subsequently modified/approved by the Subcommittee. The approved Subcommittee draft was sent to the SAB Executive Committee and their action during a publicly accessible conference call on September 11, 1998.

DISTRIBUTION LIST

INSIDE USEPA

Administrator
Deputy Administrator
Assistant Administrators
Deputy Assistant Administrator for Science, ORD
Director, Office of Science Policy, ORD
EPA Regional Administrators
EPA Laboratory Directors
EPA Headquarters Library
EPA Regional Libraries
EPA Laboratory Libraries

OUTSIDE USEPA

Congressional Research Service
Health Council of the Netherlands
Library of Congress
National Research Council: Board of Environmental Science and Toxicology
National Technical Information Service