DATE

EPA-SAB-10-00X

The Honorable Lisa P. Jackson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

Subject: Review of EPA's draft entitled "Risk and Technology Review (RTR) Risk Assessment Methodologies: For Review by the EPA's Science Advisory Board with Case Studies – MACT I Petroleum Refining Sources Portland Cement Manufacturing".

Dear Administrator Jackson:

In response to a request from EPA's Office of Air Quality Planning and Standards (OAQPS), the Science Advisory Board (SAB) convened an expert panel to review their draft document entitled, "Risk and Technology Review (RTR) Risk Assessment Methodologies: For Review by the EPA's Science Advisory Board: Case Studies – MACT I Petroleum Refining Sources Portland Cement Manufacturing" (EPA-452/R-09-006, June 2009). This draft document, hereafter referred to as the Agency's draft RTR document, describes EPA's proposed methodology for assessing residual risk from industrial emissions of hazardous air pollutants. The proposed methodologies are demonstrated through the use of two case studies: (1) petroleum refineries and (2) Portland cement manufacturing facilities. The SAB was asked to comment on seven topics addressed by the Agency's draft RTR document, including the derivation of emissions estimates, inputs for the dispersion modeling, selection of dose-response values, estimating chronic inhalation exposures, developing estimates of acute inhalation risk, developing an ecological risk assessment, and overall risk characterization.

The Panel commends the Agency on its efforts to develop a technically sound and practical approach for the challenging task of residual risk assessment. The case studies presented in the Agency's draft RTR document provide valuable insight into the strengths and limitations of the RTR inputs and methodology. While EPA proceeds with its RTR assessments, the SAB Panel recommends a number of modifications to improve the scientific basis of the RTR inputs and methodology. A more detailed description of the technical recommendations is contained in the body of the report, with the key points and recommendations highlighted.

- The Panel found emissions estimates to be one of the most critical inputs to a residual risk assessment and an important area needing improvement. As a starting point for its assessments, EPA has proposed to use "actual" emissions reported to the National Emissions Inventory (NEI), which would then be refined through internal EPA review and the public notice and comment process. However, EPA's case studies and outside evaluations suggest that the resulting emissions estimates may be biased toward underestimation. To address this concern, the Panel recommends that EPA

modify its methodology to first assess residual risks associated with facility-specific "allowable" emissions, reflecting current regulatory limits. As a second step, the Agency could model (less certain) estimates of actual facility emissions to assess current risks.

- In the particular case of radionuclides emissions from Portland cement facilities, the Panel found that information presented in the Agency's draft RTR document is not sufficiently reliable to include in the RTR assessment. Isotope-specific emissions information needs to be developed to support risk assessments for Portland cement plants and other facilities that emit radionuclides.
- The Panel found EPA's approach to selecting dose-response values to be generally sound, but recommends the Agency more closely scrutinize values that emerge as drivers of risk assessment results. The Panel supports the use of the Integrated Risk Assessment System (IRIS) as the preferred database for chronic dose-response data, and strongly recommends that EPA develop values for all HAP chemicals insofar as the data permit and update IRIS in a timely manner. The Panel recognizes that there are more gaps and inconsistencies in acute health benchmark data than in chronic data, and cautions that acute values used for residual risk assessments must be examined carefully and may need to be adjusted to ensure they protect sensitive subpopulations.
- The Panel found the ecological risk assessment case study presented in the Agency's draft RTR document to be an impressive effort to address an extremely complex issue. To further validate its RTR methodology, the Panel recommends that EPA conduct site-specific evaluations of the underlying TRIM.FaTE model and evaluate its predictions using more established ecological risk assessment methods.
- Finally, the Panel found that EPA's RTR process itself presents an incomplete picture of risks from facilities such as petroleum refineries, which fall into more than one regulatory source category. The Agency should ensure its risk characterizations clearly explain this limitation. Furthermore, the Panel agrees that RTR assessments will be most useful to decision makers and communities if results are presented in the broader context of aggregate and cumulative risks, including background concentrations and contributions from other sources in the area.

The SAB appreciates the opportunity to provide EPA with advice on this important subject. We look forward to receiving the Agency's response.

Sincerely,

Dr. Deborah L. Swackhamer, Chair
EPA Science Advisory Board

Dr. Jana Milford, Chair
SAB RTR Methods Review Panel

U.S. Environmental Protection Agency
Science Advisory Board
Risk and Technology Review Methods Panel

CHAIR

Dr. Jana Milford, Professor, Department of Mechanical Engineering, University of Colorado, Boulder, CO

MEMBERS

Dr. G. Allen Burton, Professor and Director, Cooperative Institute for Limnology and Ecosystems Research, School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI

Dr. David Eastmond, Professor and Chair, Department of Cell Biology and Neuroscience, Toxicology Graduate Program, University of California - Riverside, Riverside, CA

Mr. Thomas Gentile, Chief, Air Toxics Section, Bureau of Air Quality Analysis and Research, Division of Air Resources, New York State Department of Environmental Conservation, Albany, NY

Dr. Gary Ginsberg, Toxicologist, Environmental & Occupational Health, Connecticut Department of Public Health, Hartford, CT

Dr. Judith Graham, Independent Consultant, Independent Consultant, Pittsboro, NC

Dr. Cynthia Harris, Director and Professor, Institute of Public Health, Florida A&M University, Tallahassee, FL

Dr. Thomas W. La Point, Director, Department of Biological Sciences, Institute of Applied Sciences, University of North Texas, Denton, TX

Dr. Abby Li, Senior Science Fellow, Exponent Incorporated, San Francisco, CA,

Dr. Randy Maddalena, Scientist, Environmental Energy Technologies Division, Indoor Environment Department, Lawrence Berkeley National Laboratory, Berkeley, CA

Dr. John O'Donoghue, Adjunct Associate Professor, Department of Environmental Medicine, School of Medicine and Dentistry (Box EHSC), University of Rochester Medical Center, Rochester, NY

Dr Loren Raun, Senior Environmentalist Analyst/Faculty Lecturer, Mayor's Office, City of Houston/Rice University, Houston, TX

Dr. Mark Rood, Professor, Department of Civil and Environmental Engineering, University of Illinois, Urbana, IL

Dr. John Veranth, Research Associate Professor, Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, UT

Dr. Chris Walcek, Senior Research Scientist, Atmospheric Sciences Research Center, State University of New York, Albany, NY

SCIENCE ADVISORY BOARD STAFF

Dr. Suhair Shallal, Designated Federal Officer, 1200 Pennsylvania Avenue, NW 1400F, Washington, DC

U.S. Environmental Protection Agency
Science Advisory Board
BOARD

CHAIR

Dr. Deborah L. Swackhamer, Professor and Charles M. Denny, Jr., Chair in Science, Technology and Public Policy and Co-Director of the Water Resources Center, Hubert H. Humphrey Institute of Public Affairs, University of Minnesota, St. Paul, MN

SAB MEMBERS

Dr. David T. Allen, Professor, Department of Chemical Engineering, University of Texas, Austin, TX

Dr. Claudia Benitez-Nelson, Associate Professor, Department of Earth and Ocean Sciences and Marine Science Program, University of South Carolina, Columbia, SC

Dr. Timothy Buckley, Associate Professor and Chair, Division of Environmental Health Sciences, College of Public Health, The Ohio State University, Columbus, OH

Dr. Thomas Burke, Professor, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Dr. Deborah Cory-Slechta, Professor, Department of Environmental Medicine, School of Medicine and Dentistry, University of Rochester, Rochester, NY

Dr. Terry Daniel, Professor of Psychology and Natural Resources, Department of Psychology, School of Natural Resources, University of Arizona, Tucson, AZ

Dr. George Daston, Victor Mills Society Research Fellow, Product Safety and Regulatory Affairs, Procter & Gamble, Cincinnati, OH

Dr. Costel Denson, Managing Member, Costech Technologies, LLC, Newark, DE

Dr. Otto C. Doering III, Professor, Department of Agricultural Economics, Purdue University, W. Lafayette, IN

Dr. David A. Dzombak, Walter J. Blenko Sr. Professor , Department of Civil and Environmental Engineering, College of Engineering, Carnegie Mellon University, Pittsburgh, PA

Dr. T. Taylor Eighmy, Vice President for Research, Office of the Vice President for Research, Texas Tech University, Lubbock, TX

Dr. Elaine Faustman, Professor, Department of Environmental and Occupational Health Sciences, School of Public Health and Community Medicine, University of Washington, Seattle, WA

Dr. John P. Giesy, Professor and Canada Research Chair, Veterinary Biomedical Sciences and Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Dr. Jeffrey Griffiths, Associate Professor, Department of Public Health and Community Medicine, School of Medicine, Tufts University, Boston, MA

Dr. James K. Hammitt, Professor, Center for Risk Analysis, Harvard University, Boston, MA

Dr. Rogene Henderson, Senior Scientist Emeritus, Lovelace Respiratory Research Institute, Albuquerque, NM

Dr. Bernd Kahn, Professor Emeritus and Associate Director, Environmental Radiation Center, School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA

Dr. Agnes Kane, Professor and Chair, Department of Pathology and Laboratory Medicine, Brown University, Providence, RI

Dr. Nancy K. Kim, Senior Executive, New York State Department of Health, Troy, NY

Dr. Catherine Kling, Professor, Department of Economics, Iowa State University, Ames, IA

Dr. Kai Lee, Program Officer, Conservation and Science Program, David & Lucile Packard Foundation, Los Altos, CA

Dr. Cecil Lue-Hing, President, Cecil Lue-Hing & Assoc. Inc., Burr Ridge, IL

Dr. Floyd Malveaux, Executive Director, Merck Childhood Asthma Network, Inc., Washington, DC

Dr. Lee D. McMullen, Water Resources Practice Leader, Snyder & Associates, Inc., Ankeny, IA

Dr. Judith L. Meyer, Distinguished Research Professor Emeritus, Odum School of Ecology, University of Georgia, Lopez Island, WA

Dr. Jana Milford, Professor, Department of Mechanical Engineering, University of Colorado, Boulder, CO

Dr. Christine Moe, Eugene J. Gangarosa Professor, Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA

Dr. Eileen Murphy, Manager, Division of Water Supply, New Jersey Department of Environmental Protection, Trenton, NJ

Dr. Duncan Patten, Research Professor , Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT

Dr. Stephen Polasky, Fesler-Lampert Professor of Ecological/Environmental Economics, Department of Applied Economics, University of Minnesota, St. Paul, MN

Dr. Stephen M. Roberts, Professor, Department of Physiological Sciences, Director, Center for Environmental and Human Toxicology, University of Florida, Gainesville, FL

Dr. Amanda Rodewald, Associate Professor, School of Environment and Natural Resources, The Ohio State University, Columbus, OH

Dr. Joan B. Rose, Professor and Homer Nowlin Chair for Water Research, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI

Dr. Jonathan M. Samet, Professor and Flora L. Thornton Chair, Department of Preventive Medicine, University of Southern California, Los Angeles, CA

Dr. James Sanders, Director and Professor, Skidaway Institute of Oceanography, Savannah, GA

Dr. Jerald Schnoor, Allen S. Henry Chair Professor, Department of Civil and Environmental Engineering, Co-Director, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA

Dr. Kathleen Segerson, Professor, Department of Economics, University of Connecticut, Storrs, CT

Dr. V. Kerry Smith, W.P. Carey Professor of Economics , Department of Economics , W.P Carey School of Business , Arizona State University, Tempe, AZ

Dr. Herman Taylor, Professor, School of Medicine, University of Mississippi Medical Center, Jackson, MS

Dr. Barton H. (Buzz) Thompson, Jr., Robert E. Paradise Professor of Natural Resources Law at the Stanford Law School and Perry L. McCarty Director, Woods Institute for the Environment, Stanford University, Stanford, CA

Dr. Paige Tolbert, Associate Professor, Department of Environmental and Occupational Health, Rollins School of Public Health, Emory University, Atlanta, GA

Dr. Thomas S. Wallsten, Professor and Chair, Department of Psychology, University of Maryland, College Park, MD

Dr. Robert Watts, Professor of Mechanical Engineering Emeritus, Tulane University, Annapolis, MD

SCIENCE ADVISORY BOARD STAFF

Dr. Angela Nugent, Designated Federal Officer, 1200 Pennsylvania Avenue, NW 1400F, Washington, DC

NOTICE

This report has been written as part of the activities of the EPA 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 the 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. Reports of the EPA Science Advisory Board are posted on the EPA website at http://www.epa.gov/sab.

TABLE OF CONTENT

1.0 EXECUTIVE SUMMARY 13
1. Revisions to emissions data 14
Evaluations of petroleum refinery emissions estimates 14
Estimating dioxin and furan (D/F) emissions 16
Estimating emissions of radionuclides 17
2. Dispersion Modeling 18
3. Dose-Response Assessment 19
Selecting and prioritizing chronic dose-response values 20
Selection of acute benchmark values 21
4. Chronic Health Assessment 22
Estimating Inhalation exposures 23
TRIM model as a screening tool 23
5. Acute Health Assessment 24
6. Ecological Risk Assessment 25
7. Risk Characterization 27
2.0 Background and Introduction 29
3.0 EPA's Charge Questions 32
1. Revisions to emissions data: 32
2. Dispersion modeling: 34
3. Dose-response assessment: 35
4. Chronic health assessment: 37
5. Acute health assessment: 38
6. Ecological assessment: 39
7. Risk characterization: 40
4.0 Response to Charge Questions 41
Charge Question 1A 41
Panel Response 42
Appendix A: 44
Appendix L 45
Appendix P: 50
Charge Question 1B 54
Panel Response 55
Charge Question 1C 57
Panel Response 58
Charge Question 2 60
Panel Response 61
Use of a single year 61
Use of facility-supplied meteorology 63
Omitting atmospheric chemistry 64
Omitting deposition 65
Use of census block centroids rather than the nearest residence 66
Charge Question 3A 67
Panel Response 68
Analysis of unassessed HAPs: 70

Incorporation of HAPs lacking dose-response values: ... 71
Additional issues regarding chronic dose-response values: ... 72
Charge Question 3B ... 73
Panel Response ... 74
Adequacy of the case studies in characterizing acute risks: ... 74
Charge Question 4A ... 77
Panel Response ... 78
Charge Question 4B ... 79
Panel Response ... 80
Screening model framework and methodologies: ... 80
Communication of assumptions and results: ... 81
Omission of radionuclides from the multi-pathway assessment: ... 83
Particle bound HAPs ... 84
Charge Question 5 ... 84
Panel Response ... 85
Use of 10X annual emissions for short-term exposure estimates: ... 85
Refinements to acute screening assessment: ... 88
Summing acute hazard quotients: ... 88
Charge Question 6 ... 88
Panel Response ... 89
Facility ranking scheme: ... 93
Direct-contact ecological assessments: ... 93
Charge Question 7 ... 96
Panel Response ... 97
Presentation of risk characterizations ... 97
Risk characterization for facilities covered by more than one source category: ... 99
Characterization of aggregate and cumulative risks: ... 100
Identification and discussion of uncertainties: ... 101
Ecological risk characterization: ... 102
Cancer risk characterization: ... 103
Characterization of chronic non-cancer risks: ... 104
Characterization of acute health risks: ... 105
APPENDIX ... 106

Acronyms

ACGIH TLV	American Conference of Governmental Industrial Hygienists Threshold Limit Values
AERMET	AERMOD Meteorological Preprocessor
AERMIC	American Meteorological Society/Environmental Protection Agency Regulatory Model Improvement Committee
AERMOD	AERMIC Dispersion Model
AMS	American Meteorological Society
ATSDR MRL	Agency for Toxic Substances and Disease Registry minimum risk levels
AEGL	Acute Exposure Guidelines Limits
ANPRM	Advanced Notice of Proposed Rulemaking
CalEPA	California Environmental Protection Agency
D/F	Dioxin and Furan
ERA	Ecological Risk Assessment
ERPG	Emergency Response Planning Guidelines
HAP	Hazardous Air Pollutant
HEM-AERMOD	Human Exposure Model - AERMIC Dispersion Model
IRIS	Integrated Risk Assessment System
LOAEL	Low Observed Adverse Effect Level
MACT	Maximum Achievable Control Technology
MIR	Maximum Individual Risks
NATA	National Air Toxics Assessment
NEI	National Emissions Inventory
NESHAP	National Emission Standard for Hazardous Air Pollutants
NIST	National Institute of Standards and Testing
NOAEL	No Observed Adverse Effect Level
NPRM	Notice of Proposed Rulemaking
NWS	National Weather Service
OAQPS	Office of Air Quality Planning and Standards
OEHHA	Office of Environmental Health Hazard Assessment
PAH	Polyaromatic Hydrocarbon
PB-HAP	Persistent Bioaccumulative - Hazardous Air Pollutant
REL	Reference Exposure Levels
REM	Refineries Emissions Model
RfC	Reference Concentration
RfD	Reference Dose
RTR	Risk and Technology Review
SAB	Science Advisory Board
SMAC	Spacecraft Maximum Allowable Concentration
TRIM.FaTE model	Total Risk Integrated Methodology - Fate, Transport and Ecological Exposure Model
TRV	Toxicity Reference Values
TWE	Toxicity Weighted Emissions
UCL	Upper Confidence Limit
URE	Unit Risk Estimates
USGS	US Geological Survey

1.0 EXECUTIVE SUMMARY

This report was prepared by the Science Advisory Board (SAB) Risk and Technology (RTR) Review Panel (the "Panel") in response to a request by EPA's Office of Air Quality Planning and Standards (OAQPS) to review their draft document entitled, "Risk and Technology Review (RTR) Risk Assessment Methodologies: For Review by the EPA's Science Advisory Board: Case Studies – MACT I Petroleum Refining Sources Portland Cement Manufacturing" (EPA-452/R-09-006, June 2009). This document (hereafter referred to as the "Agency's draft RTR document") describes the Agency's proposed methodologies for conducting risk and technology review assessments demonstrated through the use of two case studies, (1) petroleum refineries and (2) Portland cement manufacturing facilities. The Panel reviewed the case studies to provide input on the RTR methodology and did not address their regulatory implications.

The Panel deliberated on the charge questions during a July 28-29, 2009 face-to-face meeting and discussed its draft report in a subsequent conference call on December 3, 2009. The final draft of the Panel's report was reviewed and approved during a meeting of the chartered SAB on XXXXXXXX. The charge questions focused on seven topics within the Agency's draft RTR document, including, the derivation of emissions estimates, inputs for the dispersion modeling, selection of dose-response values, estimating chronic inhalation exposures, developing estimates of acute inhalation risk, developing an ecological risk assessment and an overall risk characterization.

This Executive Summary highlights the Panel's major findings and recommendations resulting from their deliberations. The responses that follow represent the views of the Panel. The Panel commends the Agency on the technical quality of the draft RTR document and the thought and effort it has put into developing the residual risk methodology. The Panel found the case studies extremely valuable in illuminating both strengths and limitations of the methodology. The issues involved in residual risk estimation are extremely complex and the available information is limited. The comments and recommendations offered below are intended to assist OAQPS staff as

they seek to improve their RTR assessments going forward, and are not meant to detract from the general excellence of the Agency's draft RTR document or the efforts to date.

1. Revisions to emissions data

As described in Section 2.2.1 of the Agency's draft RTR document (*i.e.*, the Petroleum Refineries case study), the 2002 National Emissions Inventory (NEI) serves as the starting point for RTR risk assessments. EPA performs an engineering review of data from each source category to identify and correct readily apparent limitations and issues with the emissions data. The dataset is then published through an Advanced Notice of Proposed Rulemaking (ANPRM), making it available for public comment. EPA evaluates comments and corrections for quality and engineering consistency, revises the dataset, and develops a draft risk assessment. The dataset and the risk assessment are provided with a Notice of Proposed Rulemaking (NPRM) for a second 60-day comment period, after which further comments and corrections are evaluated and incorporated, as appropriate. The final rulemaking is then developed.

Evaluations of petroleum refinery emissions estimates

The Panel notes that emissions data are one of the most critical inputs to a residual risk assessment. The Panel agrees that the overall approach described in Section 2.2.1 of the Agency document is rigorous and transparent, resulting in a consistent and well documented starting point for emission scenarios. However, the Panel is concerned that the NEI, which reports *estimates* of *actual* emissions, may not be the most appropriate starting point for developing emissions data for the RTR risk assessments, due to possible underestimation bias and the potential that emissions could be increased within current regulatory limits. Where applicable, facility-specific *allowable* emissions should be considered as a first step to assess the effectiveness of the current Maximum Achievable Control Technology (MACT) standards.

Overall, the Panel found the evaluations and comparative analyses described in Appendixes A, L and P to be informative and scientifically credible. Comparisons between alternative inventory estimation methods of the maximum individual cancer risks (MIR), cancer incidence and population exposure, hazardous air pollutant (HAP) emissions, and toxicity weighted HAP emissions are useful for illustrating the key uncertainties in the current approach. However, the overarching result that emerges from the evaluations is the indication that some self-reported facility specific emissions data in the NEI are either incomplete or biased low and that the comment and revision process fails to correct this bias.

Appendix A compares risk assessment results for petroleum refineries using the emissions data from the engineering review and using emission data that were revised following the public comment period. In both cases, the analysis relies on reported emissions and does not identify or reflect further changes that may be needed to represent what MACT 1 petroleum refineries actually emit or are allowed to emit. Appendix A indicates that "facilities with a higher maximum individual cancer risk in the ANPRM were more likely to provide data changes" (p. A-8) and that these changes generally reduced the risk estimates. The Appendix would have benefited from a summary of the source of information received during the comment period to show whether comments from the facilities are generally balanced with comments from community representatives and/or state and local air pollution agencies. To ensure balanced review, the Panel recommends that EPA expand its efforts to encourage and assist community representatives to acquire relevant information and provide comments reflecting their concerns.

Appendix L compares ambient benzene concentrations with modeled concentrations for two petroleum refineries as a way to assess the emissions data. The analysis suggests the emissions data may be biased low, although inappropriate treatment of calm periods in this modeling analysis could be contributing to the apparent bias. The Panel recommends expanding the assessment to include up to 15 randomly selected refineries (~ 10 % of the total) to better represent the distribution in error across facilities. The current assessment

could also be improved by better coupling of the measurements at the source and receptor and discussing the confidence in the inventory for both facilities.

Appendix P compares risk estimates developed using NEI-based emissions inventory data with estimates developed using a process-based emissions model, the Refineries Emissions Model (REM). The comparison demonstrates differences in total emissions from refinery MACT 1 sources of a factor of almost 3 (underestimation) for benzene and a factor of 50 (overestimation) for methanol. Estimated cancer incidence for the source category is 3-4 times higher using REM emission data relative to RTR emission estimates. The Panel finds the analysis in this appendix particularly useful, as it most directly compares results based on reported NEI emissions versus estimates based on MACT compliance or "allowable" emissions.

In summary, the Panel recommends EPA modify its approach, so that facility-specific allowable emissions are modeled as a first step, to assess the effectiveness of the current MACT standards. A second step would then be modeling actual facility emissions to assess the current risk in the surrounding community. The RTR case study focuses on this second issue, but does not adequately address the concern that facilities can increase HAP emissions to MACT-allowable levels.

Estimating dioxin and furan (D/F) emissions

The Panel recommends that residual risk assessments be conducted using the current source-specific National Emission Standard for Hazardous Air Pollutants (NESHAP) allowable emission rate in combination with each facility's maximum permitted production rate. This should be done whenever NESHAP emission limits have been set for specific hazardous air pollutants. In particular, using estimated emissions that exceed the NESHAP limit is not appropriate for the residual risk assessment. Because allowable limits were not modeled for dioxin and furan (D/F) emissions from Portland cement facilities, the Panel does not believe the approach used in the case study represents the

best available methodology in support of a residual risk analysis. There is no need to estimate D/F emissions for Portland cement facilities, when allowable limits exist.

Additionally, the NESHAP compliance testing information for D/F emissions from each facility should be collected and critically evaluated to determine if it is technologically feasible to reduce the current Portland cement NESHAP D/F emission limits. This compliance information should be readily available upon request from the states or EPA regional offices. This should be done whenever NESHAP emission limits have been set for specific hazardous air pollutants. In contrast, use of the 95% Upper Confidence Limit (UCL) of available actual data as a default emission rate estimate may be appropriate for 1) source categories that do not have a NESHAP emission limit for D/F, and 2) all other HAPs that do not have a current NESHAP emission limit.

Estimating emissions of radionuclides

The Panel commends the Agency for its effort to estimate emissions and cancer risks due to radionuclide emissions from Portland cement facilities. Emissions of isotope-specific radionuclides warrant careful characterization and evaluation for Portland cement facilities and other facilities that have the potential to emit relevant radionuclides. However, the proposed analysis should not be formally included in the RTR assessment until further progress is made to quantify the isotope-specific radionuclide emissions and the associated risks. The assumptions need to be improved, as described in the body of the report, before radionuclide risk estimates are incorporated into RTR assessments. The Agency's analysis demonstrates that isotope-specific radionuclide emissions estimates are needed instead of using 2002 NEI data that do not include such speciation.

The radionuclide content of feedstocks used to produce Portland cement should be characterized at important locations across the US where these feedstocks are mined. With information on radionuclide content of feedstocks, screening material balance calculations such as those done by Leenhouts et al. (1996) for the Maastricht facility should be performed to estimate isotope-specific radionuclide emissions from Portland

cement facilities. Results from radionuclides stack tests required for compliance assurance may also provide useful data.

2. Dispersion Modeling

Section 2.2.2 describes the Agency's inputs to the AERMOD dispersion model for RTR assessments. The Agency performed these analyses in an effort to better understand the uncertainties and/or potential bias that may be introduced by some of these inputs.

The Panel believes that the dispersion modeling for primary HAPs used in risk assessments is well developed and appropriate. Any modeling entails uncertainties, and the series of case studies presented in Section 4 of the Agency's draft RTR document provide a broad picture of model performance and sensitivity for this risk assessment. The Agency has presented calculations justifying the use of several simplifications for performing longer-term impact and risk assessments. Some simplifications were shown to introduce relatively minor changes to risk estimates most of the time. However, there were some areas where simplifications introduced changes in risk estimates that could be appreciable, and in other areas further investigation is required in order to adequately justify the conclusions. In particular, it appears that there is a potentially serious underestimation bias in the dispersion modeling due to the ambiguous treatment of "calm" periods that have no definable wind directions.

The Panel noted that the choice of meteorological data for performing risk assessments appears to have a significant impact on calculated risks, as demonstrated in the sensitivity studies presented in section 4.5. The Panel also suggests that use of more than one year of meteorological observations is desirable in order to capture worst-case scenarios. The methods for choosing an individual year for risk assessment suggested here could be applied to other source categories, but depending on source stack characteristics, some of the quantitative conclusions of the Agency's sensitivity studies may not transfer.

The results of the Agency's analysis of omitting HAP *decay* and deposition in risk assessments could be applied to other source categories. However, it is possible that secondary HAP *formation* could be significant for some source categories. Further sensitivity studies of secondary HAP formation would be required to rule out the necessity of including complex photochemical modeling for future HAP risk assessments.

In order to correctly assess whether consideration of impacts at census block centroids reasonably assesses risks at actual residences within census blocks, the HEM-AERMOD system should be run twice with different sets of receptors: (1) a receptor grid of census block centroids, and (2) a receptor grid with residences tagged as receptors. Maximum health risk impacts would be directly compared using these two receptor grids for a number of facilities. It is possible that differences between block centroids and individual residences could be greater than the differences shown in this sensitivity study for source categories that are characterized by elevated buoyant emissions from smokestacks.

3. Dose-Response Assessment

Section 2.2.6 of the Agency's draft RTR document describes the process of selecting and prioritizing dose-response values for RTR human health risk assessments. The Agency selected chronic dose-response values in the same way it does for the National Air Toxics Assessment (NATA), a process that the SAB has already reviewed in the context of NATA, but not one of regulatory decision-making. The Agency has also developed an analysis (presented in Appendix O) of the possible importance of HAPs that lack chronic dose-response values. This analysis suggests that only a few HAPs lacking such values could be important in the chronic risk assessment, with the degree of importance heavily dependent on the conservatism of the input assumptions.

The Agency developed its selection process for acute dose-response values more recently than the one for chronic values, and it has not yet undergone SAB review. The acute risk

assessment process must deal with more gaps and inconsistencies in health benchmarks, compared to the chronic risk assessment.

Selecting and prioritizing chronic dose-response values

The Panel found the approach used in the RTR assessments was reasonable, but too simplistic in that it accepts dose-response numbers at face value, without closely examining the quality or validity of the value(s) chosen. In many cases, the differences in alternative chronic dose-response values will not significantly alter the RTR risk assessment, but any significant differences should be carefully considered. To assist in this effort, the Panel recommends that a table of chronic toxicity values be created, including all the chemicals under consideration and all of the eligible dose-response values, along with the source of the value, the year the value was last updated, and a qualitative description of the effect. If the chronic dose-response values are significantly different, especially if the value is a driver for the risk assessment, a review should be conducted to understand why the values differ. Professional judgment should then be used to select values for use in the assessments. If a chemical for which dose-response values have not been updated recently appears to be a driver of the assessment, a literature search should be performed to identify studies that may alter or update the value and the chemical should be considered for recommendation to the Integrated Risk Assessment System (IRIS) high priority revision list.

The preferred database for chronic dose-response data is and should be the IRIS database. However, some chemicals of interest do not have IRIS values, and values for other chemicals have not been reviewed recently. The Panel strongly recommends that the Agency address these gaps and provide the resources necessary to maintain the updating process. The use of additional sources of data should be considered; however, if additional sources of data are used they should be ones that have undergone adequate and rigorous scientific peer review.

The Panel recommends that the Agency expand the methods discussion in Appendix O to better describe the toxicity weighted emissions (TWEs) estimates for chemicals having no unit risk estimates (UREs) or reference concentrations (RfCs). In addition, the discussion of how surrogates were chosen should be clarified. Limitations about the emissions data need to be identified and addressed. The Panel recommends that the Agency prepare or compile toxicity profiles for each of the HAPs that Appendix O identifies as having the potential to drive the RTR assessment.

The issue of children's hazard should be presented as an uncertainty with regard to non-cancer dose-response assessment and carcinogen dose-response assessment - especially as only two mutagenic carcinogens receive the age-adjusted potency factor approach in the RTR, in spite of the fact that numerous other mutagens (e.g., 1,3-butadiene) are analyzed. It is not clear whether the inter-individual uncertainty factor for non-carcinogens and the standard cancer unit risk derivation adequately protect children.

The Panel was not charged with critiquing the IRIS methodology itself, however, we note below that inhalation risk methods for children are still developing and that California's Office of Environmental Health Hazard Assessment (OEHHA) has very recently updated its methodology in ways that could affect the development of RfC and URE values. EPA should examine these developments to make sure that the RTR process adequately covers children's residual risks.

Selection of acute benchmark values

The case studies characterize acute risk adequately, but this may be due to the unique circumstances of these two case studies; thus, there is a need to pay attention to the principles and practices used. The incorporation of the available California Reference Exposure Levels (RELs) for the assessment of acute effects is a conservative and acceptable approach to characterize acute risks.

The Panel does not recommend the use of the ATSDR MRLs in the risk assessments as their use would require a correction for the temporal mismatch and that correction would require formal peer-review.

The Panel has some concern with the use of the Acute Exposure Guidelines Limits (AEGLs) and Emergency Response Planning Guidelines (ERPGs). When AEGL-1/ERPG-1 emergency guideline values must be used, the Panel recommends adjusting them by a factor of 3 if the value is based on a LOAEL rather than a NOAEL. AEGL-2 and ERPG-2 values should never be used in residual risk assessments because they represent levels that if exceeded could cause serious or irreversible health effects. Spacecraft Maximum Allowable Concentrations (SMAC) for Selected Airborne Contaminants could also be considered, again with appropriate adjustments to account for the need to protect sensitive subpopulations from experiencing effects. When more reliable information is not available, American Conference of Governmental Industrial Hygienists Threshold Limit Values (ACGIH TLVs) could also be considered for use in the risk assessments, with appropriate adjustment to ensure the protection of sensitive sub-populations. TLV values should only be used after thorough and critical evaluation.

As recommended for chronic dose-response values, all the acute values for a given chemical should be arrayed in a table that displays their similarities and differences. Expert judgment should then be applied to select the most appropriate value with a clear rationale for the selection. Care must be exercised to ensure that the value chosen has undergone appropriate peer-review.

4. Chronic Health Assessment

Section 2.2.3 of the Agency's draft RTR document describes the process by which the Agency estimated chronic human inhalation exposures based on modeled average ambient concentrations at census block centroids. For these case studies, this process did not include consideration of either daily behavior pattern or long-term migration behavior. Section 2.2.3 presents a rationale for omitting daily behavior, and Appendix N

presents a case study that adjusts inhalation-based lifetime cancer risk estimates for individuals to account for long-term migration.

For persistent and bioaccumulative HAPs (PB-HAPs), the Agency's draft RTR document describes a two-step approach. As described in Appendix C, the TRIM modeling system is first used to develop what the Agency calls "*de minimis* emission rates" such that emissions below these levels should not produce unacceptable risks in reasonable worst-case conditions. Facilities emitting PB-HAPs at higher rates might require refined multi-pathway modeling, as illustrated in section 3.4 and Appendix I in a case study of a Portland cement facility.

Estimating Inhalation exposures

In general, EPA's overall approach appears to be reasonable as a *screening* approach for localized impacts that can be refined if needed in individual cases. However, an overarching concern with the Agency's chronic inhalation exposure estimates is that children's exposures do not appear to have been adequately addressed. With regard to the chronic inhalation exposure estimates, the Panel finds the rationale for omitting daily behavior to be convincing. Given the age of some of the available activity pattern data and the inherent community-scale activity pattern uncertainties between locations, the decision to omit daily behavior is justified. The Panel further recommends that long-term migration not be incorporated into the risk assessment. It does not add value to the risk assessment and introduces additional uncertainty.

TRIM model as a screening tool

In responding to this charge question, the Panel focused on how TRIM.FaTE results were applied in the risk assessment process. The Panel did not evaluate the details of the equations in TRIM.FaTE and did not itself evaluate the validity of the model. The Panel recommends that the Agency continue to identify and acquire additional field data to

estimate modeling parameters and to evaluate the TRIM.FaTE model components and other aspects of the modeling system on an ongoing basis.

With the caution that continued efforts are needed to evaluate the TRIM.FaTE model, the Panel finds that the Agency's screening approach is based on an appropriate framework and should provide a useful screen for sources that do not need a detailed site-specific multi-pathway analysis. The screening-level multi-pathway assessment is thorough and conservatively includes local subsistence agricultural and fishing scenarios, adding exposures across intake pathways to yield total PB-HAP exposure.
While the Panel supports the Agency's screening approach, we recommend EPA avoid using the term "*de minimis*" to describe the threshold emissions estimates it has derived. In particular, when the *background* concentration of a PB-HAP already exceeds a safe level (e.g., where a fish advisory is already in effect) the public may not understand a local source's contribution being characterized as *de minimis*. Furthermore, the model results should be clearly presented to show 1) the relative fraction of the local source's emissions that are deposited locally versus being transported to add to regional burdens, and 2) the relative contributions to total multi-pathway exposure from local and regional background sources.

Given the current status of information on radionuclide emissions, the Panel agrees it is acceptable to omit them from the multi-media assessment. However, EPA should work towards including them, as non-inhalation pathways are often important for radionuclides that can accumulate in biota and subsequently be ingested.

5. Acute Health Assessment

Section 2.2.5 of the draft RTR document describes the Agency's process for developing screening and refined estimates of acute inhalation risk. For acute screening purposes, the Agency has assumed that, in the worst case, a person could be exposed for one hour to ten times the highest hourly concentration calculated by the dispersion model. This in effect assumes a 1-hour emission rate of ten times (10X) the annual average (assuming

continuous emissions), simultaneous occurrence of "worst-case" meteorological conditions, and also the presence of a person at this worst-case downwind location.

The Panel agreed there is a critical need for better data addressing short-term exposures to HAPs and that in the absence of chemical- and site-specific data, the use of the 10X screening assumption for petroleum refineries seems reasonable, taking into account the aforementioned worst-case assumptions. However, the methods used to derive and justify the 10X screening assumption need to be more clearly presented. For petroleum refineries, the Panel also suggests that following the screening process, the chemicals of highest concern (drivers) be evaluated against the list of chemicals reported in the Houston area (Appendix B), to ensure they are adequately represented. Although the Panel generally agreed that the 10X assumption could be used for other geographic areas, it was felt that the actual releases would be dependent upon the manufacturing processes involved which may or may not be captured in the Houston example. As one example, adjustments may need to be made for other source categories where facilities operate during only part of the day or part of the year.

The Panel also recommends that the Agency examine the likelihood that a 10X release would occur under the most hazardous meteorological conditions, and how likely it would be for 10X releases of multiple chemicals to occur simultaneously. If it is concluded that simultaneous releases under adverse meteorological conditions would be very unlikely, then summing the acute hazard quotients by target organ would not be necessary.

6. Ecological Risk Assessment

Section 3.5 and Appendix J of the Agency's draft RTR document describe a refined, site-specific application of TRIM to conduct an ecological risk assessment for PB-HAPs emitted by the same Portland cement facility evaluated in the human health risk assessment. Appendix J also describes a nationwide facility ranking exercise that identifies Portland cement facilities with the highest potential for causing indirect

ecological effects via acidification of the environment by hydrogen chloride emissions. Appendix K describes an analysis of possible direct effects on plant foliage of air concentrations of hydrogen chloride emitted from Portland cement facilities that are below human health thresholds.

The Panel found the ecological risk assessment (ERA) presented in Appendix J to be an impressive effort tackling an extremely complex issue. While it is a good first step, the ERA needs to be improved, as the Agency's ERA guidelines were not followed well. The heavy reliance of the ERA case study on TRIM.FaTE is a concern, as the Agency's model has not been well validated in the peer-reviewed literature for ERAs, and an adequate sensitivity analysis with ground-truthing is lacking. Overall, many of the Panel's concerns and issues with the ecological risk assessment could be addressed by conducting a ground-truthing ERA at a site such as Ravena Pond, or by a comparison of TRIM.FaTE predictions with more conventional ERA methods.

The Panel recommends EPA further investigate the numerous peer-reviewed studies that are relevant to this process, many of which have focused on mercury and highly chlorinated compounds such as dioxins. In Appendix J, section 3.2.3, EPA discusses and rejects the option of using Toxicity Reference Values (TRVs) expressed in terms of tissue concentrations instead of chemical intake. However, reporting TRVs in terms of tissue concentrations (rather than intake as commonly done for human risk assessments) would allow for more and better comparisons with the peer-reviewed literature and predictions of risk, as there are fewer peer-reviewed literature reports of intake values.

The Panel found that the process to select the Portland cement facilities of greatest potential concern for HCl deposition using pH, hardness, alkalinity and soil type data was very good. However, it is important to recognize that for site-specific ERAs, other site characteristics may need to be considered

7. Risk Characterization

The risk characterizations for the two case studies (Sections 2.3 and 3.6 of the Agency's draft RTR document) represent the Agency's current practices in providing information to decision-makers responsible for RTR rulemaking. The analyses presented in the appendices are by and large illustrative of what can currently be done in the regulatory context, given knowledge, time, and resource constraints.

The Panel believes that the authors of Agency's draft RTR document took great care in summarizing and providing justification and explanation for most of the results, including attention to uncertainties. However, a number of improvements are possible. In the RTR case studies, the presentation of methods, risk assessment results, and risk characterization are intermingled, such that the purposes of the risk characterization are not met. This can be improved by focusing more on the purpose of the characterization to communicate with decision makers as the primary audience, recognizing that transparency is important and that the audience will inevitably be broad. While other sections of the RTR assessments should document the technical details, the risk characterization sections should stand alone. To this end, the Panel recommends that EPA develop a separate methods document that contains a full description (including uncertainties) of all of the common components of the source-specific risk assessments. Source-specific risk characterizations could refer back to this master document, while providing additional information particular to the source category at issue.

Decision makers and communities need to understand the broad community risk and contributors to it. However, because the Clean Air Act requires separate assessments by source category, EPA's RTR approach only partially accounts for potential human health or ecological risk of facilities that fall into more than one category. For example, the petroleum refinery MACT 1 case study omits refineries' combustion processes. The risk characterization should clearly explain this limitation. Furthermore, the risk characterization should put the results in the broader context of aggregate and cumulative risks, including background concentrations and contributions from other sources in the area.

While recognizing that RTR assessments must proceed, even though most will have a relatively long list of uncertainties, the Panel recommends that the Agency perform a sensitivity analysis to identify the major uncertainties in both the human health and ecological risk assessments. The Agency should then proceed to: (1) explain them clearly in the risk characterization section and (2) take steps to reduce them.

2.0 Background and Introduction

EPA's Office of Air Quality Planning and Standards (OAQPS) requested that the Science Advisory Board (SAB) review their draft document entitled, "Risk and Technology Review (RTR) Risk Assessment Methodologies: For Review by the EPA's Science Advisory Board: Case Studies – MACT I Petroleum Refining Sources Portland Cement Manufacturing" (EPA-452/R-09-006, June 2009). This document, hereafter referred to as the Agency's draft RTR document, describes the Agency's draft methodologies for conducting Risk and Technology Review assessments. As required by the Clean Air Act, these assessments evaluate the effects of industrial emissions of hazardous air pollutants (HAPs) on public health and the environment. The proposed methodologies are demonstrated through the use of two case studies, (1) petroleum refineries and (2) Portland cement manufacturing facilities.

The Clean Air Act establishes a two-stage regulatory process for addressing emissions of HAPs from stationary sources. In the first stage, the Act requires EPA to develop technology-based standards based on Maximum Achievable Control Technology (MACT) for categories of industrial sources. EPA must review each MACT standard at least every eight years and revise them as necessary. In the second stage of the process, EPA is required to assess the health and environmental risks that remain after MACT has been applied. EPA must develop standards to address these remaining risks if necessary to protect the public health with an ample margin of safety or to prevent adverse environmental effects. This second stage of the process is known as the residual risk review, and must be completed within eight years of promulgation of the initial MACT standards for each source category.

In order to streamline and standardize the residual risk review for the large number of source categories at issue, EPA has developed a process by which it (1) conducts a risk assessment using currently available source and emissions data; (2) requests public comment on the source and emissions data, along with preliminary risk assessment results, through an Advance Notice of Proposed Rule Making (ANPRM); (3) addresses comments received on the ANPRM; and (4) revises the risk assessment as needed. The

results of the revised risk assessment are intended to support proposals and promulgation of technology- and risk-based regulatory decisions through notice-and-comment rulemaking.

Previous SAB panels and other internal Agency and external peer review panels have reviewed aspects of the RTR methodology, as documented in the following reports:

1) The *Residual Risk Report to Congress*, a document describing the Agency's overall analytical and policy approach to setting residual risk standards, was issued to Congress in 1999 following an SAB peer review. Many of the design features of the RTR assessment methods were described in this report, although individual elements have generally been improved over the techniques described in that document. (available at: http://www.epa.gov/ttn/oarpg/t3/reports/risk_rep.pdf)

2) Individual residual risk assessments – several internal peer reviews and one external peer review were conducted on risk assessments for individual source categories, including Coke Ovens (http://www.epa.gov/ttn/atw/coke/coke_rra.pdf), Perchloroethylene Dry Cleaning (http://www.epa.gov/ttn/atw/dryperc/11-14-05riskassessment.pdf), and Halogenated Solvent Cleaners (downloadable from: http://www.epa.gov/ttn/atw/degrea/halopg.html). Each of these assessments used emission estimates from the National Emissions Inventory (NEI), human exposure modeling at the census block level, dose-response methodologies, and risk characterization that are similar to those for the planned RTR assessment.

3) The National Air Toxics Assessment, or NATA, for 1996 was peer-reviewed by an SAB panel in 2001-2002 (the SAB peer review report is available at: http://yosemite.epa.gov/sab/sabproduct.nsf/214C6E915BB04E14852570CA007A682C/$File/ecadv02001.pdf). NATA 1996 was a comprehensive and cumulative risk assessment designed to include all mobile sources, small industrial sources, and large industrial sources, as well as background contributions of air toxics. Because of significant uncertainties, the SAB did not believe that it was appropriate for

regulatory purposes. The assessment at that time did not carry a census block-level resolution, but rather was performed at the census tract level. For this reason, on EPA's NATA website (http://www.epa.gov/ttn/atw/natamain/), the estimated risks are characterized as "starting points" for developing refined assessments.

4) AERMOD, a recently-developed source-to-receptor air quality dispersion model, was the subject of significant interagency cooperation and peer review. It is now EPA's preferred local-scale air dispersion model for industrial sources of air pollution. (http://www.epa.gov/scram001/dispersion_prefrec.htm#aermod)

5) The individual dose-response assessment values used in the RTR assessment have themselves been the subject of peer reviews through the agencies that developed them (including EPA, through its Integrated Risk Information System, or IRIS; the California Environmental Protection Agency, or CalEPA, and the Agency for Toxic Substances and Disease Registry, or ATSDR). EPA proposes to select dose-response values for long-term exposures from these sources in the same priority order it used for NATA (*i.e.*, IRIS, then ATSDR, then CalEPA). For acute exposure toxicity, we array several indices without prioritization. This area is a source of significant, usually unquantifiable uncertainty. (IRIS - http://cfpub.epa.gov/ncea/iris/index.cfm, ATSDR - http://www.atsdr.cdc.gov/mrls/, CalEPA - http://www.oehha.org/air/toxic_contaminants/index.html)

6) An earlier peer review of multi-pathway risk assessment methodologies was conducted by the EPA's SAB in 2000. The final SAB advisory is available at: http://yosemite.epa.gov/sab/sabproduct.nsf/1F1893E27059DB55852571B9004730F7/$File/ecadv05.pdf.

Of particular relevance to the current review, a prior SAB panel provided a formal consultation on the *proposed* RTR Assessment methodologies in June 2007.[1] OAQPS

[1] EPA-SAB-07-009 (2007), Available at the following URL: http://yosemite.epa.gov/sab/sabproduct.nsf/02ad90b136fc21ef85256eba00436459/33152C83D29530F08525730D006C3ABF/$File/sab-07-009.pdf

revised its process to incorporate many of the SAB panel's suggestions, added significant new analysis and methods, and developed illustrative risk assessments based on the revised methodology. The current review examines the revised and expanded methodology, as illustrated through case studies for the petroleum refining and Portland cement source categories.

The Risk and Technology Review (RTR) Methods Panel met through a public teleconference call on June 30, 2009 for a briefing on EPA's Risk and Technology Review methodology and to review the charge questions presented by the Agency. The Panel then met in a public meeting on July 28 – 29, 2009 in Research Triangle Park, NC, to review the Agency's draft RTR document. The Panel held a subsequent teleconference call on December 3, 2009 to discuss its draft advisory report. The Chartered SAB conducted a quality review of this document on March 24, 2010. The responses that are contained in this report represent the views of the Panel. The specific charge questions to the Panel are as follows:

3.0 EPA's Charge Questions

EPA's charge questions for the peer review were organized into seven topic areas covering the major aspects of the proposed risk assessment methodologies. As indicated by the boxes around the text below, charge questions for several of the topic areas were grouped to focus on more specific aspects of the methodologies and case studies to which they were applied.

1. Revisions to emissions data:

As described in Section 2.2.1 of the Agency's draft RTR document (*i.e.*, the Petroleum Refineries case study), the 2002 National Emissions Inventory (NEI) serves as the starting point for RTR risk assessments. EPA performs an engineering review of data from each source category to identify and correct readily-apparent limitations and issues with the emissions data. The dataset is then published through an Advanced Notice of Proposed Rulemaking (ANPRM), making it available for public comment. EPA

evaluates comments and corrections for quality and engineering consistency, revises the dataset, and develops a draft risk assessment. The dataset and the risk assessment are provided with a Notice of Proposed Rulemaking (NPRM) for a second 60-day comment period, after which further comments and corrections are evaluated and incorporated. The final rulemaking is then developed. We have attempted to assess the quality of this process in three ways.

- Appendix A contains a comparison of risk estimates based on EPA's initial inventory as amended by engineering review and risk estimates based on the inventory as revised by public comment.

- Appendix L contains a comparison of modeled and monitored benzene concentrations around two petroleum refineries, with the intent showing if benzene emissions from refineries may have been underestimated at these facilities.

- Appendix P contains compares petroleum refinery emissions estimates and facility risk estimates using the current RTR process to emission and risk estimates from the same facilities derived using a model plant approach based on generic emission factors. The goal of this analysis was two-fold: 1) to develop a bounding estimate regarding the potential underestimation of emissions in our baseline emissions dataset; and 2) to provide an indication of how much risk estimates might change based on this potential underestimation.

1A. Do these comparisons provide useful information about the quality of the emissions data, and ultimately the risk estimates? Can you suggest improvements to these analyses, or others that might be more useful? Should we use these results to revise our risk assessment for petroleum refineries? Given that we have relatively high confidence about benzene emissions from refineries, can you suggest ways that we can develop similar analyses for other HAPs and source categories?

As described in Section 3.2.1 and Appendix F, we developed mean and upper confidence limit estimates for dioxins emitted from Portland cement facilities.

1B. Does the approach used to estimate dioxin and furan emissions from Portland cement facilities represent the best available methodology in support of a risk analysis? Can you suggest improvements?

As described in Section 3.2.2 and Appendix G, we estimated potential emissions of radionuclides, and associated inhalation cancer risks, from two Portland cement facilities using very limited data and three different derivations. The results vary by many orders of magnitude, but suggest that these risks could be substantial.

1C. Is this approach rigorous enough to consider placing it in the RTR assessment, which has regulatory implications? If not, given the lack of reliable emissions data for radionuclides, how can we improve the approach? If the quality of emissions data remains an irreducible stumbling block, can you suggest ways to obtain better emissions data?

2. Dispersion modeling:

Section 2.2.2 describes our inputs to the AERMOD dispersion model for RTR assessments. We have performed the following analyses in an attempt to better understand the uncertainties and/or potential bias that may be introduced by some of these inputs:

- Section 4.4 compares exposure estimates based on one and five years of meteorological data.

- Section 4.5 presents an analysis of how the location of the meteorological station used for modeling affects the outcome.

- Section 4.6 presents an analysis of the effect on risk estimates of omitting atmospheric chemistry from the modeling of a high-impact refinery.

- Section 4.7 presents an analysis of the effect on risk estimates of omitting deposition from the modeling of Portland cement facilities.

- Section 4.8 and Appendix M present a sensitivity analysis of the uncertainties arising in the refineries assessment by estimating exposures at census block centroids rather than at the nearest residence.

2. Do these analyses adequately support the practices of (1) using a single year of meteorological data, (2) using facility-supplied meteorological data, when available, (3) omitting atmospheric chemistry from modeling, (4) omitting deposition from modeling, and (5) using block centroids as surrogate exposure locations for these case studies? If so, can any or all of the analyses be applied to other source categories? If not, can you suggest ways we might improve them?

3. Dose-response assessment:

Section 2.2.6 of the Agency's draft RTR document describes our process of selecting and prioritizing dose-response values for RTR human health risk assessments. We select chronic dose-response values in the same way that we do for NATA, a process that the SAB has already reviewed in the context of NATA but not one of regulatory decision-making. We have also developed an analysis (presented in Appendix O) of the possible importance of HAPs that lack chronic dose-response values. This analysis suggests that only a few HAPs lacking such values could be important, with the degree of importance heavily dependent on the conservatism of the input assumptions.

3A Is our process of selecting and prioritizing chronic dose-response values appropriate for RTR risk assessments? Should we consider additional sources, or a different

prioritization process? Can the analysis of unassessed HAPs be improved by developing prior assumptions regarding the toxicity of these HAPs, and if so, how should this be done? Are there other ways we can improve it? Is this approach inherently limited to the current bounding exercise and tool for identifying research needs, or can it be further developed and incorporated into RTR assessments? Can you provide advice on how we can incorporate HAPs lacking dose-response values into our risk characterizations?

We developed our selection process for acute dose-response values more recently than the one for chronic values, and it has not yet undergone SAB review. The universe of acute health benchmarks contains many gaps, as shown in Table 2-5. In addition, some of the benchmarks correspond to "no-effect" levels (*e.g.*, CalEPA acute reference exposure levels, which are analogous to chronic RfCs), while others correspond to "mild-effect" or "severe-effect" levels (*e.g.*, acute exposure guideline levels) that are intended to guide authorities in making emergency evacuation decisions. For these reasons we have not applied a prioritization scheme.

We have not generally included acute minimum risk levels (MRLs, developed by the Agency for Toxic Substances and Disease Registry, or ATSDR) as dose-response values in our assessments of acute risks because of a temporal mismatch between the exposure estimates (based on one hour) and the MRLs (based on 24 hours to two weeks).

3B. Given these gaps and inconsistencies among available acute benchmarks, do the case studies characterize acute risks adequately? Should we include ATSDR MRLs in our assessments, and if so, how can we solve the temporal mismatch? Is the use of emergency guidelines in our assessments adequately described and interpreted? Are there other acute health metrics EPA should consider using for these assessments? Do you have suggestions for improvements in any of these areas?

4. Chronic health assessment:

Section 2.2.3 describes the process by which we estimate chronic human inhalation exposures based on modeled average ambient concentrations at census block centroids. For these case studies, this process did not include consideration of either daily behavior pattern or long-term migration behavior. Section 2.2.3 presents a rationale for omitting daily behavior, and Appendix N presents a case study that adjusts inhalation-based lifetime cancer risk estimates for individuals to account for long-term migration.

4A. Does our process of estimating inhalation exposures adequately support regulatory rulemaking? Is our rationale for omitting daily behavior convincing, or does the omission compromise the value of our assessments? Should this, or some other, adjustment for long-term migration be incorporated into our risk assessments?

Appendix C describes a novel application of TRIM in the development of protective *de minimis* emission rates for 14 persistent and bioaccumulative HAPs (PB-HAPs). We believe that emissions below *de minimis* thresholds should not produce unacceptable risks in reasonable worst-case conditions. Facilities emitting below these values would not need to conduct a multi-pathway exposure and risk assessment.

Section 3.4 and Appendix I describe a refined application of the TRIM model in assessing multi-pathway pollutant transport and its subsequent impacts on human health from Portland cement facility air emissions identified as having a high potential to present significant impacts on human health.

We have limited our development of radionuclide risk estimates (described in Section 3.2.2 and Appendix G) to those associated with inhalation exposure. Radionuclides were not included in the multi-pathway risk assessment.

4B. Is our use of the TRIM model to develop *de minimis* emission rates appropriate as a screening tool? Are the methodologies used in the refined multi-pathway assessment

consistent with the best available science regarding multi-pathway pollutant transport and human exposures? Are the application of the model and the assumptions used clearly articulated? Are the resultant estimates of media concentrations and exposures clearly presented, explained, and interpreted? Given the large uncertainties surrounding the radionuclide inhalation assessment, are we justified in omitting radionuclides from the multi-pathway assessment?

5. Acute health assessment:

Section 2.2.5 describes our process for developing screening and refined estimates of acute inhalation risk. For acute screening purposes we have assumed that, in the worst case, a person could be exposed for one hour to ten times the highest hourly concentration calculated by the dispersion model. This in effect assumes a 1-hour emission rate of ten times the annual average (assuming continuous emissions), simultaneous occurrence of "worst-case" meteorological conditions, and also the presence of a person at this worst-case downwind location.

Appendix B presents an effort to evaluate the protectiveness of this screening assumption using detailed short-term emission data for a limited geographic area. Appendix E describes our refinement of acute risk estimates for refineries that failed the acute 10X screen, by using more accurate emission points and property boundaries.

Our refined acute assessments do not combine acute hazard quotients associated with different HAPs because of the inconsistent nature of acute health benchmarks and the inherent conservatism of our exposure assumptions.

5. Does the 10X acute screening assumption for petroleum refineries appear to be appropriately protective? If not, is it under- or over-protective? Given that this analysis applies only to sources in the Houston area, can we apply the 10X assumption to HAPs in other source categories or should we consider some other approach for some other HAPs, *e.g.*, metals? Is there some other way we might address high emission events such as

startup or shutdown of processes? Are the refinements to the acute screening assessment objectively employed and scientifically defensible? Should we sum acute hazard quotients by target organ in the same way we do for chronic hazard quotients, *i.e.*, a target organ specific hazard index (TOSHI) approach, or are our reasons for not doing so adequate?

6. Ecological assessment:

Section 3.5 and Appendix J describe a refined, site-specific application of TRIM to conduct an ecological risk assessment for PB-HAPs emitted by the same Portland cement facility evaluated in the human health risk assessment. Appendix J also describes a nationwide facility ranking exercise that identifies Portland cement facilities with the highest potential for causing indirect ecological effects via acidification of the environment by hydrogen chloride emissions. Appendix K describes an analysis of possible direct effects on plant foliage of air concentrations of hydrogen chloride emitted from Portland cement facilities that are below human health thresholds.

6. Is the ecological assessment case study scientifically defensible? Does it conform to EPA risk assessment guidance (*e.g.*, *Guidelines for Ecological Risk Assessment*, *Risk Characterization Handbook*, *etc.*)? If not, how can we improve it? Are the elements of the ranking scheme adequate to identify the facilities most likely to be of concern? Are there better data sources or approaches for drawing conclusions for specific locations? With regard to investigating the potential for direct ecological effects at air concentrations below human health thresholds from other sources or source categories, what suggestions can be made for prioritizing additional HAPs for literature searches similar to that done for hydrogen chloride in Appendix K?

7. Risk characterization:

The risk characterizations for these two case studies (Sections 2.3 and 3.6) represent our current practices in providing information to decision-makers responsible for RTR rulemaking. The analyses presented in the appendices are by and large illustrative of what can currently be done in the regulatory context, given time and resource constraints.

7. Do these characterizations objectively and completely incorporate the goals and principles of EPA's *Risk Characterization Handbook* to the extent scientifically feasible? In particular do they provide a complete and transparent discussion of uncertainties and limitations? If not, how can the risk characterizations be improved? Can you suggest where we might focus any additional efforts and resources in order to have the biggest impact on refining risk characterizations for these RTR assessments, ultimately leading to better regulatory decision-making?

4.0 Response to Charge Questions

Charge Question 1

As described in Section 2.2.1 of the Agency's draft RTR document (*i.e.*, the Petroleum Refineries case study), the 2002 National Emissions Inventory (NEI) serves as the starting point for RTR risk assessments. EPA performs an engineering review of data from each source category to identify and correct readily-apparent limitations and issues with the emissions data. The dataset is then published through an Advanced Notice of Proposed Rulemaking (ANPRM), making it available for public comment. EPA evaluates comments and corrections for quality and engineering consistency, revises the dataset, and develops a draft risk assessment. The dataset and the risk assessment are provided with a Notice of Proposed Rulemaking (NPRM) for a second 60-day comment period, after which further comments and corrections are evaluated and incorporated. The final rulemaking is then developed. We have attempted to assess the quality of this process in three ways.

- Appendix A contains a comparison of risk estimates based on EPA's initial inventory as amended by engineering review and risk estimates based on the inventory as revised by public comment.

- Appendix L contains a comparison of modeled and monitored benzene concentrations around two petroleum refineries, with the intent showing if benzene emissions from refineries may have been underestimated at these facilities.

- Appendix P contains compares petroleum refinery emissions estimates and facility risk estimates using the current RTR process to emission and risk estimates from the same facilities derived using a model plant approach based on generic emission factors. The goal of this analysis was two-fold: 1) to develop a

bounding estimate regarding the potential underestimation of emissions in our baseline emissions dataset; and 2) to provide an indication of how much risk estimates might change based on this potential underestimation.

1A. Do these comparisons provide useful information about the quality of the emissions data, and ultimately the risk estimates? Can you suggest improvements to these analyses, or others that might be more useful? Should we use these results to revise our risk assessment for petroleum refineries? Given that we have relatively high confidence about benzene emissions from refineries, can you suggest ways that we can develop similar analyses for other HAPs and source categories?

Panel Response

Emissions data are one of the most critical inputs to a residual risk assessment. The process for deriving emission factors for the risk and technology review (RTR) risk assessments begins with the 2002 National Emissions Inventory (NEI) data compiled for individual facilities in a given source category. The data are reviewed and revised by EPA (engineering review) followed by a two-stage public comment process (ANPRM and NPRM) leading to further revisions in response to comments. EPA has invested a great deal of effort into adapting and applying the existing NEI data to construct emissions scenarios for the RTR assessments. The Panel agrees that the overall approach described in Section 2.2.1 of the Agency's draft RTR document is rigorous and transparent, resulting in a consistent and well documented starting point for emission scenarios based on an existing and well documented data set. However, the Panel is concerned that the NEI data, which reports *estimates* of *actual* emissions, may not be the most appropriate starting point for developing emissions data for the RTR risk assessments, due to possible underestimation bias and the potential that emissions could be increased within current regulatory limits. Where applicable, the Panel recommends that facility-specific *allowable* emissions be considered as a first step, to assess the effectiveness of the current MACT standards.

EPA performed three modeling analyses for the petroleum refineries case study to assess the quality of the process for developing RTR emissions data. The first analysis (Appendix A) compares the outcome of the risk assessment using emissions data from before and after the comment period to explore how the public comment process influenced the outcome. The second (Appendix L) compares modeling results for benzene concentrations to monitoring results at two facilities to determine if emission factors may be underestimated. The third (Appendix P) compared the current approach to a category specific emissions modeling approach using generic emission factors to explore the potential for underestimation of emissions in the base-line scenario and how this might influence risk estimates.

Overall, the Panel found the analyses described in Appendixes A, L and P to be informative and scientifically credible. Comparisons in the analyses such as the maximum individual cancer risks (MIR), cancer incidence and population exposure, HAP emissions, and toxicity weighted HAP emissions are useful for illustrating the key uncertainties in the current approach. However, the overarching result that emerges from the evaluations is the indication that some self-reported facility specific emissions data in the NEI are either incomplete or biased low and that the comment and revision process fails to correct this bias.[2]

It is the Panel's understanding that the Agency is aware of the deficiencies in the petroleum refineries emission estimates. The City of Houston recently submitted a request for correction of information under the Data Quality Act and EPA's Data quality guidelines[3]. The request cites reports of underestimation of emissions by up to two orders of magnitude for refineries and chemical manufacturing plants. The EPA responded in a letter[4] dated April 7, 2009, expressing concurrence with the City's concerns and acknowledging the inaccuracy and uncertainty of emission estimates in the inventory, particularly where there is heavy reliance on emission factors in the NEI. The

[2] See also Document ID EPA-HQ-OAR-2003-0146-0010, "Potential Low Bias of Reported VOC Emissions from the Petroleum Refining Industry", EPA Technical Memorandum from Brenda Shine, July 27, 2007.
[3] http://www.greenhoustontx.gov/reports/epaletter20080709.pdf
[4] http://www.greenhoustontx.gov/reports/dataquality20090407.pdf

Agency outlined a number of specific tasks that are currently on-going to address and fully understand this uncertainty. The planned outcome of this work, as described by the EPA, is to use the results of the emission factor verification project to help: a) evaluate risk to exposed populations; b) conduct comparisons to existing emission estimates (e.g. TRI) for specific facilities; and c) better characterize the cost effectiveness of controls. The Panel is concerned that any residual risk decision made for the petroleum refinery source category without the use of this updated and verified emissions information would be premature.

The Panel's review of the appendices is discussed below followed by recommendations for improving the emission estimates for the RTR process.

Appendix A: The stated purpose of Appendix A is to compare the risk assessment results using the emissions data from the engineering review with results using revised emission data that were revised following the public comment period. In addition to changes in the emissions data, a number of other changes were made to the risk assessment between the two cases. For example, Appendix A indicates that although the same risk assessment model (HEM3/AERMOD) was used in both assessments, several updates were made in the version used with the post-comment emissions data. Specifically, the meteorological data included additional meteorological stations and a newer version of the AERMET model was used along with meteorological data from different (more recent) years. In addition, updated dose-response data were used for the post-comment assessment. The appendix is silent on the potential impact of these changes relative to changes in the emissions data. Although it is likely that emissions are the dominant factor influencing the changes in the results, the validity of this assumption is not demonstrated.

The comparison in Appendix A is focused on reported actual emissions. Thus the assessment does not identify or reflect further changes that may be needed to represent what MACT 1 petroleum refineries actually emit (as opposed to what they report emitting) or what they might emit if emissions were increased to allowable levels under

existing MACT standards. The analysis would be more informative if it included adjustments to the HAP emissions from all facilities needed to reflect representative emissions across the source category.

Another important observation from Appendix A is the relationship between the likelihood of receiving input during the public comment period and the magnitude of the individual risk values reported in the ANPRM. Figure 6 of Appendix A highlights the fact that comments were more likely to be provided for facilities for which individual cancer risk was relatively high and that these comments generally reduced the risk estimates. There is clear incentive for facilities associated with higher risk to offer corrections to the NEI data but it is unclear whether similar incentives are present to help identify underreporting facilities. The analysis would have benefited from a summary of the source of information received during the comment period to evaluate whether the comments originating from groups representing the facilities are generally balanced with comments from groups representing the community, or if facility-specific emissions data were submitted by state and local air pollution agencies. In many cases, community representatives might not have the expertise or access to emissions information to provide substantive input to the review process. Most state and local air pollution agencies rely on the emission factors contained in EPA's AP-42 Fifth Edition Compilation of Air Pollutant Emission Factors, Volume 1: Stationary Point and Area Sources[5] to estimate facility emissions unless they have facility specific emissions testing data. To ensure balanced review, the Panel recommends that EPA expand its efforts to encourage and assist community representatives to acquire relevant information and provide comments reflecting their concerns.

Appendix L: The Panel recognizes that evaluating model performance using empirical observations is very important for increasing confidence in model-based assessments. In this Appendix, ambient benzene concentrations measured at two sampling locations were compared to modeled concentrations at or near the same sample locations for two facilities as a way of assessing the emissions data used in the risk assessment at these

[5] AP-42 Fifth Edition Compilation of Air Pollutant Emission Factors, Volume 1: Stationary Point and Area Sources. Available On-Line: http://www.epa.gov/ttnchie1/ap42/

facilities. The assessment assumes that the dominant source of variation in modeled concentrations at the sample locations is the emissions data used in the model runs. The Appendix shows that modeled concentrations are significantly lower than monitored concentrations with the difference for one facility (Marathon facility) being much greater than the other. The Agency's draft RTR document points out that a statistically significant difference does not necessarily imply practical importance. However, the analysis clearly shows both an apparent low bias in the emissions data and a low precision in the predictions from the two facilities. The analysis thus suggests the emissions data may be biased low, although inappropriate treatment of calm periods in this modeling analysis could be contributing to the apparent bias.

While the model results suggest that emissions are biased low, it is notable that the results for the two facilities are very different. Annual averaged modeled concentrations are within 11% of the corresponding monitored values for the BP facility, but only within 72% for the Marathon facility. Correspondingly, the absolute errors between the measured and modeled annual average concentrations are 0.5 $\mu g/m^3$ and 3.4 $\mu g/m^3$ for the two petroleum refineries. Given that the $1x10^{-6}$ cancer risk benchmark for benzene is an annual average concentration of 0.128 $\mu g/m^3$, the absolute error is considerable. The difference in error between the two refineries highlights the problem with using a small sample size (n=2 out of 154 refineries) to assess model performance. The small and co-located sample of two facilities makes it difficult to conclude that a high level of confidence exists in the evaluation of benzene emissions based on these results. Furthermore, the analysis depends on extensive assumptions about averaging of emissions, characterizing surface roughness, and characterizing the meteorology. The comments offered by the internal EPA reviewer about difficulties in characterizing wind speed and direction closer to the receptors, and not including emissions from additional sources (e.g., ship/barge traffic) are appropriate and may limit the value of this assessment.

Monitored ambient concentrations represent the sum of contributions from all sources. In order to estimate the portion of the ambient concentration that could be attributed to the

source category or the specific facility, EPA used the following general methodology: 1) monitors in close proximity to the source were used, 2) data were evaluated by wind direction so that it could reasonably be assumed that concentrations at the monitor were related to the source (when the monitor was downwind from the source) and, 3) concentrations not attributed to the source (e.g., on-road mobile, background estimated from NATA) were subtracted from the total concentration seen at the monitor. The background estimate appears to be a type of correction factor and an attempt to account for the contributions from other sources. It is unclear where the background estimate came from and/or if dispersion modeling was used to derive it. It may be that the background value is actually a crude combination of unaccounted for fugitive emissions and error from under reporting in the emission inventory. Because the background may reflect errors in reported emissions, it may not be appropriate to subtract this source off hand from the ambient concentrations. While it is important to account for background given the long half-life of benzene, the analysis should provide a better description of the background estimate, including where it comes from and its spatial distribution.

The choice of using meteorology from the more distant site (Galveston) when local information was available seems incorrect. Ordinarily the closest meteorological monitor should be utilized. The model-to-data comparison in this appendix needs to be appropriately adjusted under the assumption that a potentially significant error could have been introduced into these comparisons by using incorrect meteorology. The fact that there is general agreement of the plume positioning with wind direction suggests that the winds in Galveston statistically resemble the winds further inland at the refinery location, but hour-by-hour discrepancies may be significant. Although clustering of sites that behave in a similar manner is seen farther up the ship channel, the Galveston airport site is likely to act more independently given its location. This site is open and closer to the Gulf. Uncertainty in the wind direction and speed could be brought into the model and spatially assessed. EPA defends its use of data from the Galveston airport site by pointing out questions in the representativeness of the Texas City Ball Park site, which is closer to the refineries. Re-evaluation which includes a margin of error is the only way to ascertain the influence of the issues with the wind data. An additional examination of the model to

monitor comparison for these two facilities using the closer meteorological data set would be useful.

The assessment could also be improved by better coupling of the measurements at the source and receptor and discussing the confidence in the inventory for both facilities. This would strengthen the analyses. From the background documentation contained in the Air Docket (EPA-HQ-OAR-2003-0146), it appears that the BP-Texas City facility has provided a credible assessment of their inventory based on the limited model to monitor comparisons and the findings of the 22 facility study that indicated BP-Texas City seemed to properly account for benzene emissions from their storage tank facilities in comparison to other facilities.[6] The confidence in the inventories for the two facilities could also be discussed in light of other findings from the 22 facility study, such as the finding that many facilities underestimate their benzene emissions from the wastewater stream by as much as a factor of 40 to 1400.

The Panel recommends expanding the assessment to include up to 15 randomly selected refineries (~ 10 % of the total) to better represent the distribution in error across facilities. It is unlikely that the discrepancy between reported and actual emissions can be assumed to be constant between facilities. To gain a better understanding of the modeled to measured error, a stratified random sample of refineries assigning strata based on, for example: size of the facility (our experience suggests that large facilities, even those that are well run, tend to have more fugitive emissions error and more error in general simply from having more sources); age of the facility (older facilities may not operate as well); compliance record (facilities with more violations may have larger under-reporting error than other facilities). It appears there may be a more robust dataset (more benzene ambient air monitors located near petroleum refineries) that could be assembled and evaluated in a more comprehensive manner.

If the analysis is limited by available monitoring data, the Panel recommends that rather than using a strict comparison of the model and monitoring results, the two data sets

[6] Lucas, Bob. (2007). Technical Memorandum to EPA Docket No. EPA-HQ-2003-0146 from Bob Lucas, EPA/SPPD dated August 20, 2007.

might be used in conjunction to provide a more comprehensive understanding of the probability or range of outcomes using, for example, a Bayesian approach. At a minimum, it would be useful to include a more formal uncertainty analysis and consider propagation of errors to better quantify the uncertainties and characterize the agreement with the benzene concentration data (See Bevington's book "Data Reduction and Error Analysis"[7]).

Finally, Appendix L attempts to put the potential error into context of the overall errors expected in the risk assessment, but may be misleading in this regard. The statement on page L-1 regarding the analysis of the measured to modeled concentrations says,

> "[The analysis] attempts to answer the question, "are benzene emission estimates truly lower by a factor of 10 to 100 (at least for these 2 facilities), or are they close enough to be useful in residual risk decision-making?" We attempt to answer this last part keeping in mind the 2 orders of magnitude range of MIR values embodied in the residual risk decision framework."

This statement is not very clear, but could be interpreted to mean that the Agency might not view the level of uncertainty resulting from emissions estimates as a large concern, given that the risk range for risk management decisions under the Clean Air Act spans two orders of magnitude. But such a view could be misleading. Even if less than a factor of 10, an *underestimation bias* in the emissions estimates should still raise concerns, as it could prevent a source category from falling into the residual risk range that would otherwise require remedial action. In contrast, as discussed below, questions such as whether the centroid of a census block is modeled or population migration is included may be on a level of detail and sophistication rendered obsolete given the inherent uncertainty of the emissions input data.

[7] Bevington, Philip R. and D. K. Robinson. Data Reduction and Error Analysis for the Physical Sciences, New York: McGraw-Hill Companies, 2003.

Appendix P: This appendix compares risk estimates developed using RTR emissions inventory data with estimates developed using emissions estimates from a process-based emissions model, the Refineries Emissions Model (REM). The results are informative. The comparison demonstrates differences in total emissions from refinery MACT 1 sources (Appendix P, Table 1) of a factor of almost 3 for benzene and a factor of 50 for methanol. However, xylenes and POM 72002 are in agreement to within about 50%. There is a wide range in the ratio of REM MIR value to RTR MIR value for individual facilities (p. P-23), ranging from 0.1 to 5,000,000 (with all but one value ranging from 0.1 to 5,000). Also, estimated cancer incidence for the source category is 3-4 times higher using REM emission data relative to RTR emission estimates (Appendix P, Table 3). Instructive comparisons are also provided for specific emission sources such as fugitive equipment leaks, cooling towers, HAP storage vessels, and areas for wastewater collection and treatment.

The assessment illustrates the problem the EPA encountered with the development of the emissions inventory for this source category. The analysis in this appendix actually almost addressed the Panel's concerns about the use of actual emissions as reported in the 2002 NEI. It states that the modeled REM emissions are based on MACT compliance or allowable emissions. The difference shown in Table 3 between the RTR-estimated "actual" HAP emissions (6,820 tons/year) and the REM allowable HAP emissions (17,800 tons/year) that are known to be emitted by MACT 1 petroleum refineries is stark. It is difficult to compare the risk results between these two emissions estimates and agree with the conclusion that the REM database results in a "modest increase in risk estimates" for the following reasons:

(1) the RTR used site specific emission point data (18 to 42% of the time) to estimate community impacts while the REM used default emission source release parameters for all HAP emissions placed them in the centroid of petroleum refining facilities and then estimated the risk at the centroid of the census block. This approach can underestimate the resultant MIR risk. The impact of consolidating emissions points into a centroid emissions point for large facilities

with multiple emissions points has been found to underestimate impacts in the area close to the facility property boundary by a factor of 3 to 7[8];

(2) the emissions estimates change the MIR cancer risk drivers (REM drivers are benzene, naphthalene and POM as compared to the RTR drivers naphthalene and POM);

(3) the REM-based analysis excludes two more toxic groups of POM that would result in an increase of the MIR and cancer incidence;

(4) the REM analysis results in increases in the cancer incidence and MIR ranking of the facilities even though the two more toxic groups of POM are excluded; and

(5) neither the RTR nor the REM emissions inventories attempt to account for emission releases due to upsets and malfunctions.

The Panel does not agree with the closing statement of Appendix P, "Petroleum Refineries are highly regulated facilities for which emissions are thought to be relatively *well understood* (emphasis added) as compared to many other source categories. The relative similarity in MIRs may be unique in this case. It is difficult to generalize the results of this analysis to other source categories". This Panel is concerned this statement may convey a false degree of confidence in the emissions inventory that is not warranted for the source category as a whole, based on the information provided in the case studies.

<u>Recommendations related to Charge 1A</u>: The comparisons provided in Appendices A, L, and P provide a transparent and useful look at the quality of the available emissions data for use in the RTR assessments. However, the results do not instill a high degree of confidence about the hazardous air pollutant (HAP) emissions inventory. The HAP emissions inventory is the foundation of the residual risk assessment. A poor emissions inventory will result in a poor residual risk assessment. The underestimation of emissions will result in false negatives or underestimation of community risk while the overestimation of emissions and reporting of HAPs that are not expected to be emitted

[8] USEPA, 1998. Analysis Performed for the Risk Screening Environmental Indicators. Office of Pollution Prevention and Toxics. Available On-Line: http://www.epa.gov/oppt/rsei/pubs/index.html

from the source category will result in false positives or overestimation of community risk.

It is readily apparent that the quality of the facility-specific HAP emissions inventories ranges from good to poor. Table 2-6 (p. 2-22) clearly illustrates this problem. There are 156 facilities in this data set and they do not consistently report emissions that are expected for MACT 1 petroleum refinery processes. For example, only 146 out of 156 facilities report benzene emissions, 129 facilities report xylene emissions, 136 facilities report toluene emissions, 130 facilities report hexane emissions and 104 report naphthalene emissions. There is no consistent reporting of polycyclic organic matter (POM) across facilities, although POM is one of the identified RTR cancer risk drivers. There are emissions of polyaromatic hydrocarbons (PAHs) total, POM, 16-PAH and individual PAHs by the facilities. It is unclear how any meaningful risk analysis could be undertaken for these emissions. There are five facilities that report a total of three tons of carbon tetrachloride emissions. The production and use of this HAP has been banned under the 1990 Clean Air Act. While there are expected to be regional differences for some HAPs emitted from this source category (i.e. methanol and MTBE), some HAPs (e.g., benzene, xylene, toluene, and hexane) should be reported by all facilities in the source category.[9]

The RTR case study models actual emissions using the 2002 National Emissions Inventory (NEI) and there apparently was an adjustment of these emissions using site-specific data from 22 refineries as provided by the American Petroleum Institute. However, it is not clear what adjustments were made. In particular, it is not clear whether all of the facilities' emissions inventories were adjusted by using the information contained in the August 6, 2007 technical memorandum on the Average Refinery Stream Composition. This technical memorandum clearly identifies the product specific HAP emissions that should be expected from the sources subject to the MACT 1 Petroleum Refineries NESHAP.

[9] Lucas, Bob. (2007). Technical Memorandum to EPA Docket No. EPA-HQ-2003-0146 from Bob Lucas, EPA/SPPD dated August 6, 2007.

The Panel recommends that EPA classify the emissions inventory (actual emissions) for the 156 facilities subject to this MACT standard by simple degrees of confidence (high, medium or low). The categorization of the 156 facilities should consider size, throughput capacity and product refined. This evaluation should also include statements about the confidence in the AP-42 emission factors for the source category. The AP-42 manual already has a ranking system for all of the individual chemical emission factors. So a characterization of the confidence in these values for the specific process emissions under evaluation should be included in the residual risk assessment.

The primary goal of the residual risk assessment should be to assess the impacts of HAPs in the surrounding community within the bounds of what is permissible or allowable by the National Emission Standard for Hazardous Air Pollutants. As a first step, the facility-specific MACT 1 allowable emissions should be modeled. The modeling of NESHAP or MACT allowable emissions is necessary since the individual facilities are allowed by federal regulation to emit HAPs in these quantities into the surrounding community. The EPA cannot accurately assess effectiveness of the NESHAP to reduce risk and be protective of public health and the environment by modeling actual emissions from these facilities, especially if the actual emissions are way below what is allowed to be emitted by the NESHAP. Beyond modeling residual risk from allowable emissions, a second step would be the modeling of actual facility emissions to assess the current risk in the surrounding community. The RTR case study focuses on this second issue, but it does not adequately address the issue that these facilities can increase HAP emissions to permissible NESHAP levels.

The Panel recommends that EPA model REM allowable emissions using the same emissions point information and toxicity factors as used in the RTR to properly assess the residual risk associated with sources regulated by the National Emission Standards for Hazardous Air Pollutants from Petroleum Refineries (Code of Federal Regulations Part 63 Subpart CC) (MACT1 Petroleum Refineries). This type of analysis will better assist EPA to meet with greater confidence the two-fold goal of the RTR as stated in the June 17, 2009 memorandum containing the charge questions to the SAB.

The Panel has some additional suggestions for improving the HAP emissions inventory for these and other source categories subject to residual risk assessments. First, EPA could adopt a consolidated emissions reporting rule for hazardous air pollutants that requires all major facilities subject to Part 63 NESHAPs to uniformly report their actual and allowable emissions along with emission point parameters on an annual or semi-annual basis. The two case studies presented in this review and previous residual risk assessments appear to have suffered because of the lack of a federal requirement to report HAP emissions in a consistent and uniform manner. An alternative way to address this issue is to rely on facility specific compliance inspection information (state and federal) and Section 114 data requests. The information collected during compliance and enforcement proceedings is some of the most thorough information collected on facility specific emissions. Unfortunately, these data are usually sealed until an enforcement action is completed and in most cases will reflect sources that are out of compliance with state and federal air pollution standards. The mining of these data is also labor intensive. A third alternative would be to work closely with state and local air pollution control agencies to gather any facility specific emissions testing data that can be useful in the preparation of residual risk assessments.

Charge Question 1B

As described in Section 3.2.1 and Appendix F, we developed mean and upper confidence limit estimates for dioxins emitted from Portland cement facilities.

1B Does the approach used to estimate dioxin and furan emissions from Portland cement facilities represent the best available methodology in support of a risk analysis? Can you suggest improvements?

Panel Response

The primary purpose of the risk and technology review (RTR) for Portland cement facilities is two-fold: (1) to evaluate the residual risk to public health and the environment that remains after the application of the initial technology or emission limits contained in the Portland cement NESHAP; and (2) to critically analyze the performance of the air pollution control requirements of the current NESHAP and evaluate whether the original allowable dioxin/furan (D/F) emission limits could be reduced further, if this is shown to be technologically feasible by actual testing data. For the first step of this process, the Panel recommends that residual risk assessments be conducted using the current source-specific NESHAP allowable emission rate in combination with each facility's maximum permitted production rate. This should be done whenever NESHAP emission limits have been set for specific hazardous air pollutants. In particular, using estimated emissions that exceed the NESHAP limit is not appropriate for the residual risk assessment. Because allowable limits were not modeled for D/F emissions from Portland cement facilities, we do not believe the approach used in the case study represents the best available methodology in support of a residual risk analysis. There is no need to estimate D/F emissions for Portland cement facilities, when allowable limits exist.

The final Portland Cement NESHAP, 40CFR Part 63 LLL contains two D/F emission limits: (i) 0.20 nanograms per dry standard cubic meter (8.7×10^{-11} grains per dry standard cubic foot) (TEQ); or (ii) 0.40 nanograms per dry standard cubic meter (1.7×10^{-10} grains per dry standard cubic foot) (TEQ) when the average of the performance test run average temperatures at the inlet to the particulate matter control device is 204 °C (400 °F) or less. For new and existing Portland cement kilns, the residual risk assessment should model these currently allowable emission rates of D/F in combination with stack flow rates corresponding to maximum permitted production rates for each facility. The information needed for this assessment should be available from the required compliance testing information for every Portland cement facility identified in the case study. If these allowable D/F emission limits result in an unacceptable risk to public health and the

environment after the completion of the multi-pathway risk assessment as conducted in the case study, a decision to lower these existing D/F limits should be made.

It appears that if EPA used allowable D/F emissions in its analysis, none of the Portland cement facilities considered would screen out of needing a refined multi-pathway assessment based on the emission thresholds presented in Appendix C-4.5.1. However, since the risk from D/F exposure is primarily driven by the fish and beef/dairy consumption exposure pathways, EPA could consider screening out facilities that have negligible potential to impact fishable waters and beef and dairy farms.

In the second step of the RTR process, the NESHAP compliance testing information for D/F emissions from each facility should be collected and critically evaluated to determine if it is technologically feasible to reduce the current Portland cement NESHAP D/F emission limits. This compliance information should be readily available upon request from the states or EPA regional offices. The information presented in the case study demonstrates that the D/F emissions from the various kiln types can significantly vary. The review of actual compliance data by kiln type could lead to the establishment of lower D/F emission limits by kiln type sub-categorization as determined through a technology review of the existing compliance data. The review should also address the issue that many Portland cement kilns burn alternative fuels that are not classified as hazardous waste (tire-derived fuel, used oil) and the influence of these materials on dioxin emissions needs to be considered and noted in any future analyses. The availability of the D/F compliance testing data for this source category should result in a more robust analysis of the technological feasibility of lowering these D/F limits by kiln type, which is independent of the residual risk assessment requirement.

A specific comment about how the risk assessment information for D/F is presented in Portland cement case study is warranted. The Agency should be cognizant of how the results of the residual risk assessments will be perceived by the public in the impacted communities. Public concerns about the impacts of D/F emissions are extremely high. The methodology used in the case study could raise unnecessary public concern about fish consumption in the community, the consumption of beef and dairy produced in the

surrounding area, and adverse effects on wildlife that would not be warranted if the Ravena plant is in compliance with the current NESHAP D/F emission limit. Based on additional information that EPA provided to the SAB Panel, the use of the 95% UCL emission factor developed for wet kilns and listed in Table F-3 would result in a violation of the current NESHAP D/F emission limit. The application of this emission factor in the residual risk assessment would result in a false positive risk result or an overestimate of the MIR risk. In general, residual risk assessments should rely on the use of NESHAP allowable emission rates when available for specific hazardous air pollutants in combination with maximum production rates. In contrast, use of the 95% UCL of available actual data as a default emission rate estimate may be appropriate for i) source categories that do not have a NESHAP emission limit for D/F, and ii) all other HAPs that do not have a current NESHAP emission limit.

Finally, EPA needs to carefully verify the emission point parameters it uses in its analysis for the Portland cement industry. The stack exit temperature they used in the case study for the Ravena facility appears to be off by 115 ^{o}F. The Agency's draft RTR document lists it as 350°F whereas the 2003 stack testing report for the facility indicates it is 465 ^{o}F.

Charge Question 1C

As described in Section 3.2.2 and Appendix G, we estimated potential emissions of radionuclides, and associated inhalation cancer risks, from two Portland cement facilities using very limited data and three different derivations. The results vary by many orders of magnitude, but suggest that these risks could be substantial.

1C Is this approach rigorous enough to consider placing it in the RTR assessment, which has regulatory implications? If not, given the lack of reliable emissions data for radionuclides, how can we improve the approach? If the quality of emissions data remains an irreducible stumbling block, can you suggest ways to obtain better emissions data?

Panel Response

The Panel commends EPA for its effort to estimate emissions and cancer risks due to radionuclide emissions from Portland cement facilities. Emissions of isotope-specific radionuclides warrant careful characterization and evaluation for Portland cement facilities and other facilities that have the potential to emit relevant radionuclides.

EPA's proposed approaches to estimating inhalation cancer risks due to radionuclide emissions from Portland cement facilities indicate that such risks could be substantial. EPA found more than 80 of the 91 facilities assessed had estimated Maximum Incremental Risks (MIR) from radionuclide releases in excess of 2×10^{-6} (Exhibit G-12). However, the proposed analysis should not be formally included in the RTR assessment until further progress is made to quantify the isotope-specific radionuclide emissions and the associated risks. The revised approach should also consider the potential for multi-pathway exposure of isotope specific radionuclides (e.g., dietary exposure pathways, Exhibit G-13).

The Agency's draft RTR document relies heavily on non-isotope specific radionuclide emissions reported in the 2002 National Emissions Inventory (NEI) for two Portland cement facilities in California and on results from emission modeling for radionuclides at the Maastricht Portland cement facility in the Netherlands (Leehouts et al., 1996, http://rivm.openrepository.com/rivm/bitstream/10029/10172/1/610053003.pdf). EPA provided alternative evaluations with emissions estimated by scaling ^{210}Po and ^{222}Rn to clinker production; scaling to particulate matter (PM) emissions; and by assuming all radionuclide emissions reported to the NEI were either ^{210}Po or ^{222}Rn. EPA clearly stated the assumptions used in estimating the radionuclide emissions under each approach. However, the assumptions need to be improved as described below before radionuclide risk estimates are incorporated into RTR assessments.

EPA's analysis demonstrates that isotope-specific radionuclide emissions estimates are needed instead of using 2002 NEI data that do not include such speciation. In particular, emissions and risk estimates EPA obtained by assuming NEI radionuclide mass

emissions were all ^{210}Po were implausible, illustrating the importance of completing careful engineering review of input data before beginning risk modeling.

Radionuclides such as uranium and thorium also exist in many geological materials at ppm(m) concentrations. The radionuclide content of feedstocks used to produce Portland cement should be characterized at important locations across the US where these feedstocks are mined. Other toxic trace elements, such as mercury, could also be considered at the same time. Such information should be available in the literature, as it is for other geologic materials such as fossil fuels. EPA's Indoor Environments Division (IED, located within ORIA and under OAR), the US Geological Survey (e.g., Radioactive Elements in Coal and Fly Ash: Abundance, Forms, and Environmental Significance, USGS Fact Sheet FS-163-97, Oct 1997), the National Institute of Standards and Testing (NIST), the US Nuclear Regulatory Commission, and nuclear engineering and geology departments at academic institutions are possible sources of such information. Any source category that has the potential to cause increased local exposure to airborne radon and polonium needs to have this issue addressed as part of the RTR process.

With information on radionuclide content of feedstocks, screening material balance calculations such as those done by Leenhouts et al. (1996) for the Maastricht facility should be performed to estimate isotope-specific radionuclide emissions from Portland cement facilities. This analysis should use data for US feedstocks and estimate the atmospheric emissions that would occur after implementing MACT. Thus, a much improved screening for potential radionuclide emissions should be performed by using mean and upper confidence limit literature data for isotopes in the feed materials and information about the operating conditions of the facility (e.g., temperature and chemical reactions in the process). Such information may also provide insight as to how to reduce radionuclide emissions during the production of Portland cement.

If results from revised screening calculations are not acceptable or data are not available to support such analyses, then source information describing isotope-specific

radioactivity should be obtained from select Portland cement facilities, including results from stack tests. Such information should include descriptions of the isotope-specific radionuclides that are processed and then emitted from the Portland cement facilities.

Emission characterization of the radionuclides could also be improved by evaluating closure between measured radioactivity at receptors near a Portland cement facility to radioactivity predicted using estimated source strengths and dispersion modeling; this evaluation would be similar to what was done for the petroleum refinery case study in this review. The feasibility of undertaking such an evaluation assumes ambient radioactivity levels are detectable near the sources, considering background values and detection limits of analytical techniques.

Charge Question 2

Section 2.2.2 describes our inputs to the AERMOD dispersion model for RTR assessments. We have performed the following analyses in an attempt to better understand the uncertainties and/or potential bias that may be introduced by some of these inputs:

- Section 4.4 compares exposure estimates based on one and five years of meteorological data.

- Section 4.5 presents an analysis of how the location of the meteorological station used for modeling affects the outcome.

- Section 4.6 presents an analysis of the effect on risk estimates of omitting atmospheric chemistry from the modeling of a high-impact refinery.

- Section 4.7 presents an analysis of the effect on risk estimates of omitting deposition from the modeling of Portland cement facilities.

- Section 4.8 and Appendix M present a sensitivity analysis of the uncertainties arising in the refineries assessment by estimating exposures at census block centroids rather than at the nearest residence.

2 Do these analyses adequately support the practices of (1) using a single year of meteorological data, (2) using facility-supplied meteorological data, when available, (3) omitting atmospheric chemistry from modeling, (4) omitting deposition from modeling, and (5) using block centroids as surrogate exposure locations for these case studies? If so, can any or all of the analyses be applied to other source categories? If not, can you suggest ways we might improve them?

Panel Response

The dispersion modeling for primary HAPs used in risk assessments is well developed and appropriate. Any modeling entails uncertainties, and the series of case studies presented in Section 4 provide a broad picture of model performance and sensitivity for this risk assessment. EPA has presented calculations justifying the use of several simplifications (i.e., assumptions) for performing longer-term impact and risk assessments. Some simplifications were shown to introduce relatively minor changes to risk estimates most of the time. However, there were some areas where simplifications introduced changes in risk estimates that could be appreciable, and in other areas further investigation is required in order to adequately justify the conclusions. The following discussion highlights some of the impacts of these assumptions on the risk assessment.

Use of a single year of meteorology: The sensitivity analysis of the use of one versus five years of meteorological observations is well done, and shows that most of the time, uncertainties of less than 10% are introduced in calculated concentrations, although maximum annual or hourly concentrations can differ by up to 10-40% at some locations and times. While the conclusion of this section suggests that uncertainties in risk estimates due to the inclusion of more meteorological observations are minor if reported risk estimates are limited to one significant figure, we suggest that use of more than one

year of meteorological observations is desirable in order to capture worst-case scenarios. At most sites, numerous years of meteorology observations are available and should be examined to ensure impacts are not underestimated.

If more meteorological observations are used in any longer-term impact analysis, markedly higher concentrations and impacts may be encountered on hourly scales, while annual averages are expected to fluctuate by smaller amounts (relative to maximum hourly impacts) under the influence of more smoothly-varying averaged year-to-year meteorological variations. It is standard EPA procedure in New Source Review permitting to utilize five years of meteorological data, and the SAB recommends following this protocol when feasible. Unless there are serious computational or labor resource limitations, we suggest that maximum annual-average impacts be defined from the worst year of several years' analysis. Acute impacts should be calculated using the worst 1-hr impacts calculated using whatever number of years of meteorological data is available for analysis.

It appears that there is a potentially serious underestimation bias in the dispersion modeling due to the ambiguous treatment of "calm" periods that have no definable wind directions. This factor could be contributing to AERMOD calculating lower concentrations than observed, as seen in the petroleum refineries case study (Appendix L). The highest concentrations generally occur during calm periods, and the emissions modeling analysis appears to ignore calm periods, treating them as equivalent to missing meteorological measurements. By ignoring these periods, potentially significant errors that underestimate maximum concentrations will result. Such a simplification needs to be investigated before concluding that emissions information might be biased low. EPA should clarify how calm periods are treated in AERMOD, and consider whether the approach needs to be revised to avoid underestimating risks and health impacts.

The methods for choosing an individual year for risk assessment suggested in the Agency's document could be applied to other source categories, but depending on source stack characteristics, some of the quantitative conclusions of EPA's sensitivity studies

may not transfer. Apparently for the refinery source category used in this 1 versus 5 year sensitivity study, HAP emissions were mostly ground-level sources without significant stack heights or plume rise. For other source categories that are emitted in buoyant plumes or from elevated stacks, the confounding effects of plume rise will appreciably influence calculated impacts, and it is possible that differences between 1 and 5 years of meteorology could be greater than the differences shown in this sensitivity study, which was dominated by ground-level sources.

Use of facility-supplied meteorology[10]**:** The choice of meteorological data for performing risk assessments appears to have a significant impact on calculated risks, as demonstrated in the sensitivity studies presented in section 4.5. In this section, EPA compared risk estimates for four petroleum refineries that were derived using meteorological data from three to five different meteorological stations, each within about 200 km of the source. The "overall summary" of this section that "differences usually fall within rounding error for the one-significant-figure characterization of risk" is somewhat inconsistent with the results shown in Table 4-2, which show that differences greater than a factor of two are common, and there is no consistent trend in these differences with distance from emission source. In all likelihood, these appreciable differences result from the fact that even the closest National Weather Service (NWS) meteorological monitoring station only crudely captures the hourly meteorology that is representative of conditions near emission sources and impact receptors. Over broad areas, especially in the western U.S., there can be gross errors introduced in air quality impacts calculated using the closest NWS meteorological monitoring station. Sometimes several mountain ridges or valleys may lie between a particular site and a meteorological monitor. Given the small horizontal scales of 1-hr winds (boundary-layer scale – less than 1-2 km), one would expect discrepancies similar to those shown in this sensitivity study for monitors separated by only 1-10 km from some source locations. As noted in

[10] It appears that this charge question is poorly worded, since the use of "facility provided" meteorology is not addressed in the sensitivity study. In the preamble to the charge question it is noted that the study covers the "location of meteorological station", and section 4.5 mentions that two refineries furnished meteorology data, but results from these "facility-supplied" meteorology are not presented. Therefore, the Panel interpreted the charge question to more generally consider EPA's selection of meteorological stations.

comments on Appendix L, it would be desirable to use facility-provided meteorology for risk assessments, if available. Unfortunately, site-specific meteorology is probably not available for most facilities, and this remains a significant source of uncertainty in any risk assessment calculation. The potential errors introduced by using meteorology that is not representative of a given source or receptor location is partially ameliorated by using as long a record of meteorological data as is computationally feasible, to increase the probability that high impact conditions are encountered and included.

The underlying problem of using meteorology that is representative of each source location is an endemic problem for any risk assessment irrespective of the source category considered. The best method for quantifying whether the closest NWS station meteorology is "representative" of any emission source point would be to quantitatively compare source-specific onsite meteorology measurements with nearby NWS monitors, and perform sensitivity analysis comparing the use of onsite meteorology versus using the nearest NWS observations as input. This direct comparison was not done in this analysis, as onsite data were not included in the comparisons, so the conclusions of this section suggesting that using "nearest NWS" site meteorology introduces relatively minor uncertainties in risk assessments is not well established by the sensitivity studies presented here.

Omitting atmospheric chemistry: Many emitted HAPs undergo relatively slow photochemical oxidation following release. The sensitivity study presented in section 4.6 addresses only the decrease in concentrations of emitted (primary) HAPs due to oxidation during photochemical aging. It is well known that the time scales for photochemical transformations of most HAPs are considerably longer than the transport times between sources and highly impacted receptors, and therefore the concentrations of emitted HAPs will decrease by relatively small amounts due to photochemical processes. Under these conditions, ignoring atmospheric chemistry would be reasonable for these risk assessments, and the sensitivity study presented in section 4.6 adequately demonstrates this.

However, several organic HAPs (e.g., formaldehyde) are formed during the oxidation of other emitted volatile organic compounds, and it is not obvious that ignoring photochemical formation of secondary (formed) HAPs is reasonable. Therefore, an additional study of secondary HAP *formation* needs to be performed in order to rule out the need for incorporating complex photochemistry in these risk assessments. Such a sensitivity study could involve running a short-term (2-4 hour simulation) photochemical "box model" including a gas-phase chemical mechanism under typical daytime conditions for a broad range of VOC/NOx emission profiles representative of various source categories, then estimating the secondary formation of HAPs such as formaldehyde. The calculated concentrations of secondary HAPs from a simple box model alone could provide concentrations that could then be used as inputs to screening models of potential risk assessments to ascertain whether secondary HAP formation could be an important contributor to air quality risk endpoints.

The results of EPA's analysis of the omission of HAP *decay* in risk assessments could be applied to other source categories. However, it is possible that secondary HAP *formation* could be significant for some HAP source categories. As noted above, further sensitivity studies of secondary HAP formation would be required to rule out the necessity of including complex photochemical modeling for future HAP risk assessments.

Omitting deposition: It is well known that the time scales for deposition are considerably longer than the transport times between sources and highly impacted receptor locations, so during this time the concentrations of emitted HAPs will decrease by relatively small amounts due to deposition. Under these conditions, ignoring deposition would be reasonable for these risk assessments. Section 4.7 confirms this conclusion through a rigorous and reasonable comparison of risk assessments performed with and without deposition, showing changes of a few percent or less for a handful of facilities. Based on this study, it is expected that the simplification of omitting deposition could be generalized to other source categories when performing HAP risk assessments.

Use of census block centroids rather than the nearest residence: This analysis suggests that cancer risks calculated at census block centroids are usually the same, or sometimes considerably greater than (up to 2000%!) risks calculated at individual residences within a census block. This analysis appears to contain some fundamental simplifications that render the results somewhat ambiguous. It appears that risk impacts have been interpolated to residence locations from centroid and polar grid receptors, rather than explicitly calculated using AERMOD (Appendix M). Furthermore, the residence impacts have been unrealistically set to centroid impacts if the census blocks are "small", or if residences are "near" the centroid, or if the polar grid was "not adequate" to interpolate to a particular residence. These vague interpolation methods will produce residence impacts that are identical to the centroid impacts quite often in an unrealistic fashion. It is also possible that the conclusions of this sensitivity study may be an artifact of the particular configurations of census block maps and residence locations used for the subset of facilities (21 of 154) chosen. It is possible that large underestimations of risk could occur for other facilities, other source categories, or census block/residence configurations.

In order to determine whether impacts at census block centroids reasonably assess risks at actual residences within census blocks, the HEM-AERMOD system should be run twice with different sets of receptors: (1) a receptor grid of census block centroids, and (2) a receptor grid with residences tagged as receptors. Maximum health risk impacts would be directly compared using these two receptor grids for a number of facilities. The AERMOD model itself should be run for actual residences in order to accurately assess risks at those residences.

Another area of concern related to this sensitivity study entails the use of a limited subset (21 of 154) of facilities considered. In order to compare centroid versus residence impacts and draw general conclusions, it is not necessary to explicitly simulate all 154 facilities associated with this source category; a carefully chosen, stratified subset of facilities could be used to draw more general conclusions. In this study, the subset was restricted to the 21 facilities with the greatest MIR. These 21 facilities may not be representative of

the range of possible census-block/ residence locations, meteorology, and source configurations that would influence the differences between impacts at residences and centroids. Clearly a better criterion must be used to define a "representative" subset of test cases. For example: urban, suburban and rural facility locations should probably be sampled, even if some of these facilities have low impacts.

It is possible that the conclusions of any sensitivity study of receptor locations will not be generally applicable to other source categories. HAP emissions for this sensitivity study are dominated by ground-level sources without significant stack heights or plume rise. Under these conditions the greatest impacts will be in census blocks closest to the facilities. For other source categories that are emitted in buoyant elevated stacks, the confounding effects of plume rise can move the regions of greatest impact further from the source locations. It is possible that differences between block centroids and individual residences could be greater than the differences shown in this sensitivity study for source categories that are characterized by elevated buoyant emissions from smokestacks.

Charge Question 3A

Section 2.2.6 of the Agency's draft RTR document describes our process of selecting and prioritizing dose-response values for RTR human health risk assessments. We select chronic dose-response values in the same way that we do for NATA, a process that the SAB has already reviewed in the context of NATA but not one of regulatory decision-making. We have also developed an analysis (presented in Appendix O) of the possible importance of HAPs that lack chronic dose-response values. This analysis suggests that only a few HAPs lacking such values could be important, with the degree of importance heavily dependent on the conservatism of the input assumptions.

3A Is our process of selecting and prioritizing chronic dose-response values appropriate for RTR risk assessments? Should we consider additional sources, or a different prioritization process? Can the analysis of unassessed HAPs be improved by developing prior assumptions regarding the toxicity of these HAPs, and if so, how should this be

done? Are there other ways we can improve it? Is this approach inherently limited to the current bounding exercise and tool for identifying research needs, or can it be further developed and incorporated into RTR assessments? Can you provide advice on how we can incorporate HAPs lacking dose-response values into our risk characterizations?

Panel Response

Process of selecting and prioritizing chronic dose-response values**:** The approach used in the RTR assessments is reasonable, but too simplistic in that it accepts dose-response numbers at face value, without much understanding of the quality or validity of the value(s) chosen. Of concern is that some values have been developed quite some time ago using older data, which may be obsolete, while others have been developed more recently and incorporate new findings. Even dose-response values that use the same up-to-date database are not equivalent, as different agencies do not derive hazard values in the same way. For example, for the benchmark methods, EPA and CalEPA apparently both take the lower 95th confidence limit of the dose of interest, but then look at the dose level that causes a 10% (EPA) or a 1 or 5% (CalEPA) incidence of the critical effect. In many cases, the differences in chronic dose-response values will not significantly alter the RTR risk assessment, but they do suggest a need to carefully consider any significant differences in chronic dose-response values so that the credibility of the risk assessment is not impaired by selection of an outdated data point.

To address this concern, the Panel recommends that a table of chronic toxicity values be created. The table should include all the chemicals under consideration, all of the eligible dose-response values (e.g., if both EPA and CalEPA have values for the same chemical, both should be included), the source of the value, the year the value was last updated, and a qualitative description of the effect (e.g., eye irritant, neurotoxicant, reproductive toxicant, cancer classification) as all effects do not have equal health impacts. The entries in the table should be reviewed for consistencies among the values available for each chemical. If the chronic dose-response values are significantly different between agencies, especially if the value is a driver for the risk assessment, a review should be

conducted to understand why the values differ. By necessity, professional judgment will need to be used during the chronic dose-response value selection process to decide which value is most appropriate to use based upon thoroughness of the data review, consistency of the dose-response modeling with the underlying science base, and the Agency's objectives for public health protection. All of this analysis can be part of an appendix, with the text only having the information selected for use in the assessment.

Furthermore, if a chemical appears to be a driver of the assessment, the assessor should further review the value and examine how recently it had been developed. If it was developed more than 7 years ago, a literature search should be performed to identify studies that may alter or update the value. If such studies are identified, the chemical should be considered for recommendation to the Integrated Risk Assessment System (IRIS) high priority revision list for review of the dose-response value.

The preferred database for chronic dose-response data should be the IRIS database. The Panel strongly recommends that EPA update the values in IRIS and provide the resources necessary to maintain the updating process. Concern about the quality of the IRIS database and approaches to keeping it up-to-date have previously been addressed by the SAB and others.[11] The Panel endorses these recommendations for change in the IRIS database and process for updating the database.

The use of additional sources of data should be considered; however, if additional sources of data are used they should be ones that have undergone adequate and rigorous scientific peer review. The inclusion of additional sources of dose-response values into the hierarchy needs to be adequately documented in a transparent manner in any residual risk assessment case study.

[11] As stated on page 56 of the Residual Risk Report to Congress under the heading, Data Availability, Limitations, and Closing Data Gaps, the preferred source of dose-response data for conducting federal risk assessments is the IRIS database. However as discussed in a recent GAO report (available at http://www.gao.gov/new.items/d09773t.pdf), the IRIS database is at serious risk of becoming obsolete due to an absence of timely updates of existing IRIS values and a significant backlog of ongoing assessments.

The American Conference of Governmental Industrial Hygienists Threshold Limit Values (ACGIH TLVs®) could be considered for use as an additional source of data for screening purposes, when other values are not available. The TLVs have been determined for healthy workers; therefore, for use in the residual risk process, the values would require time adjustment from a 40-hr workweek to a 24 hr/day, 7 day/ week exposure (168 hrs/week). Further adjustment for consideration of protection for susceptible populations would be needed, and if a TLV is not considered a No Observed Adverse Effect Level (NOAEL), another adjustment factor might be needed.

Analysis of unassessed HAPs: The SAB has previously commented on the importance of having reliable dose-response values for all of the HAPs listed in the 1990 Clean Air Act.[12] The residual risk exercise emphasizes, once again, the importance of having accurate, current information in the Agency's IRIS database.

Appendix O provides the rationale for selecting dose-response values based on chemicals that have already been thoroughly evaluated. It is an interesting attempt to fill the void and create some type of toxicity ranking scheme to prioritize HAPs for toxicity testing and dose-response assessment and for the use of surrogate reference concentrations (RfC) and unit risk estimate (URE) values in the residual risk assessment process. There appears to be extremely limited and highly variable information about the emissions of some of these HAPs, which handicaps the prioritization process. The HAPs that are being reviewed by this process have large data gaps for which professional judgment is needed to derive surrogate RfCs and UREs. This approach creates more uncertainty in the selection of a surrogate RfC or URE for use in the residual risk case studies.

We assume based on our reading of the case study text that surrogates were chosen as follows: All values in Table 1 of the indicated reference were evaluated for percentiles, resulting in the table at the top of page O-2. Thus, a chemical having no URE or RfC is assumed to fall into the same percentiles as chemicals that had such values. Then the

[12] Review of the US EPA's report to Congress on Residual Risk. EPA-SAB-EC-98-013; Advisory on the USEPA's draft Case Study Analysis of the Residual Risk of Secondary Lead Smelters. EPA-SAB-EC-ADV-00-005; Advisory from the National-scale Air Toxics Assessment. NATA – Evaluating the National-Scale Air Toxics Assessment 1996 Data – SAB Advisory. EPA-SAB-EC-ADV-02-001.

emissions of a chemical having no URE or RfC were multiplied by the percentiles, creating values that show up on Figure O-1. The Panel recommends that the Agency expand the methods discussion to include a better description of the toxicity weighted emissions (TWEs) for chemicals having UREs and RfCs, using some of the language from the Air Toxics Risk Assessment Reference Library (see http://www.epa.gov/ttn/fera/data/risk/vol_3/Appendix_B_April_2006.pdf or http://www.epa.gov/ttn/fera/data/risk/vol_1/chapter_06.pdf).

In addition, the discussion of how surrogates were chosen should be made clearer. Limitations about the emissions data need to be identified and addressed. For example, only one facility out of 104 Portland cement facilities reports 48 tons per year of carbonyl sulfide. This questionable emissions data drives the TWE process in Appendix O and carbonyl sulfide is listed as a priority HAP for further dose response evaluations. If our assumptions above about the calculations of surrogates are correct, and a verification of the emissions inventory is conducted the approach is adequate, if limited to screening purposes.

Any unassessed HAPs that screen-in because of this evaluation process should then be followed-up by reviewing existing toxicity information to examine the likelihood that they could be a driver for the assessment process.

The current bounding exercise and tool for identifying research needs is limited to this purpose and probably cannot be further developed and incorporated in the RTR assessments given the limitations of the emissions inventory for these HAPs**.** HAP-specific emissions testing would have to be conducted at these facilities in order to use and have confidence in weighting factors that are based on the amount of actual HAPs released.

Incorporation of HAPs lacking dose-response values: The Panel recommends that the Agency prepare or compile toxicity profiles for each of the HAPs that Appendix O identifies as having the potential to drive the RTR assessment. They should receive a

very high priority for evaluation according to the IRIS process that was recently redefined by Administrator Lisa Jackson. (See http://www.gao.gov/new.items/d09773t.pdf for a review of recommendations and changes to be made to the IRIS process). Residual risk decisions for these chemicals will have to be identified as awaiting peer review or Agency-wide consensus.

Additional issues regarding chronic dose-response values: The Panel was not charged with critiquing the IRIS methodology itself and therefore was not constituted with the expertise for an in-depth review of the methodology. However, we note below that inhalation risk methods for children are still developing and that California Office of Environmental Health Hazard Assessment (OEHHA) has very recently updated its methodology in ways that could affect the development of RfC and URE values. US EPA should examine these developments to make sure that the RTR process adequately covers children's residual risks.

In particular is the question of whether the interindividual variability factor for non-carcinogens and the standard cancer unit risk derivation adequately covers children. If it does not, it is a potentially significant uncertainty given the greater intake rate of children via inhalation and sensitivity to carcinogens and other toxicants.[13]

California EPA/OEHHA has determined that inhalation dosimetry for children is sufficiently different from adults to warrant a full 10-fold intra-individual pharmacokinetic uncertainty factor (i.e., an extra 3-fold PK uncertainty for children relative to the IRIS method) as a default approach. In setting non-cancer reference exposure levels (RELs), Cal EPA/OEHHA also considers that children may be outliers in terms of chemical susceptibility and on a case-specific basis adds a children's

[13] USEPA, 2005. Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens EPA/630/R-03/003; Barton HA, Cogliano J,Flowers L, Valcovic L, Setzer RW, Woodruff TJ. 2005. Assessing Susceptibility from Early-Life Exposure to Carcinogens. Environ Health Perspectives 113:1125-1133; Hattis D,Goble R,Russ A,Chu M, Ericson J. 2004. Age-related differences in susceptibility to carcinogenesis: a quantitative analysis of empirical animal bioassay data. Environ Health Perspectives 112:1152-1158.

pharmacodynamic factor of 3-fold, making the inhalation risk for children as much as 10 times greater than adults).[14]

This issue of childrens's hazard should be presented as an uncertainty with regard to non-cancer dose-response assessment and carcinogen dose-response assessment, especially considering that only two mutagenic carcinogens receive the age-adjusted potency factor approach in the RTR, even though numerous other mutagens (e.g., 1,3-butadiene) are analyzed. California's OEHHA uses the children's cancer potency adjustment factors on a much broader array of carcinogens than the narrow interpretation used in the Agency's draft RTR document.[15] This would be a natural area for sensitivity analysis (e.g., applying the age-adjusted potency factor to numerous carcinogens (at least all those that are mutagens) to determine the degree of uncertainty children's vulnerability can create in the cancer risk assessment.

Charge Question 3B

We developed our selection process for acute dose-response values more recently than the one for chronic values, and it has not yet undergone SAB review. The universe of acute health benchmarks contains many gaps, as shown in Table 2-5. In addition, some of the benchmarks correspond to "no-effect" levels (*e.g.*, CalEPA acute reference exposure levels, which are analogous to chronic RfCs), while others correspond to "mild-effect" or "severe-effect" levels (*e.g.*, acute exposure guideline levels) that are intended to guide authorities in making emergency evacuation decisions. For these reasons we have not applied a prioritization scheme.

We have not generally included acute minimum risk levels (MRLs, developed by the Agency for Toxic Substances and Disease Registry, or ATSDR) as dose-response values in our assessments of acute risks because of a temporal mismatch between the exposure estimates (based on one hour) and the MRLs (based on 24 hours to two weeks).

[14] (http://www.oehha.ca.gov/air/hot_spots/2008/NoncancerTSD_final.pdf
[15] http://www.oehha.ca.gov/air/hot_spots/tsd052909.html

3B Given these gaps and inconsistencies among available acute benchmarks, do the case studies characterize acute risks adequately? Should we include ATSDR MRLs in our assessments, and if so, how can we solve the temporal mismatch? Is the use of emergency guidelines in our assessments adequately described and interpreted? Are there other acute health metrics EPA should consider using for these assessments? Do you have suggestions for improvements in any of these areas?

Panel Response

Adequacy of the case studies in characterizing acute risks: The case studies characterize acute risk adequately, but this may be due to the unique circumstances of these two case studies. Thus, there is a need to pay attention to the principles and practices used. The incorporation of the available California Reference Exposure Levels (RELs) for the assessment of acute effects is a conservative and acceptable approach to characterize acute risks.

The Panel has some concern with the use of the Acute Exposure Guidelines Limits (AEGLs) and Emergency Response Planning Guidelines (ERPGs). These limits were developed for accidental release emergency planning and are not appropriate for residual risk assessments without modification because, as described in the AEGL and ERPG documentation, adverse effects may occur at these levels. For example, at the AEGL-1 level, "…the general population, including susceptible individuals, could experience notable discomfort, irritation, or certain asymptomatic nonsensory effects. However, the effects are not disabling and are transient and reversible upon cessation of exposure." (citation-in each AEGL document). Some of the AEGLs and ERPGs listed in Table 2-5 are higher than values used to protect healthy workers from acute effects in occupational settings. The Panel recommends considering reducing the AEGL-1/EPRG-1 emergency guideline values by a factor of 3, when the value is based on a LOAEL rather than a NOAEL. This would better approximate a "no-effect" level, as in RfC's. In contrast,

AEGL-2 and ERPG-2 values should never be used in residual risk assessments because they were derived on the basis of maximum concentrations that would result in serious or irreversible health effects if they were exceeded.

The short-term exposure levels (STELs) and ceiling levels used by the American Conference of Governmental Industrial Hygienists (ACGIH), Occupational Safety and Health Administration (OSHA) and the National Institute of Occupational Safety and Health (NIOSH) were developed to protect healthy workers from short exposures that may routinely occur in the workplace. The use of acute dose response values that are greater than occupational values used to protect healthy workers does not provide a high degree of confidence that the dose response values used in the case studies have adequately characterized the acute risk of HAP exposures for sensitive subpopulations within a community. For example, the use of the AEGL-1 for 1,3-butadiene (1500 mg/m^3) versus the OSHA short-term exposure limit (11 mg/m^3) calls into question the adequacy of the use of emergency planning values in any residual risk assessment.

The Panel does not recommend the use of the ATSDR MRLs in the risk assessments as their use would require a potentially complex correction for the temporal mismatch. In order to use the MRL values, the risk assessors would have to recalculate an acute value based on the critical endpoint(s) identified in the ATSDR documentation. Appropriate safety factors would have to be determined and applied to the critical endpoint to determine an acceptable acute exposure value. Without peer review of the calculated value, the credibility of the assessment would be questionable even in a screening assessment.

When other more reliable values are not available, it is recommended that adjusted occupational values (ACGIH TLV) be considered for use in the risk assessments. The acute TLV values represent an evaluation of the literature that, by using expert judgment, could be adjusted and considered for use. Because substantial risk may remain with

exposure at TLV levels, these values should only be used after thorough and critical evaluation.[16]

Other sources of peer-reviewed health values are the Spacecraft Maximum Allowable Concentrations for Selected Airborne Contaminants (SMACS).[17] SMACS are defined as "the maximum concentrations of airborne substances that will not produce adverse health effects, cause significant discomfort, or degrade crew performance" and are classified into 1- and 24-hour emergency SMACS and 7-, 30-, and 180-d continuous SMACS. SMACS are developed in a similar way to other health values, except that they typically do not include an uncertainty factor for susceptible subpopulations because the target population is a healthy adult population. Furthermore, many of the SMACS represent effect levels, rather than "safe" levels, so they would need to be dealt with in a manner similar to emergency values. It is recommended that EPA add these documents to its list of sources for analysis. Because susceptibilities are not accounted for, these values would need to be divided by an uncertainty factor of 3 or 10 (similar to the adjustment recommended for AEGL-1 or other acute values), and then compared to other values. Also, if the SMAC was related to a LOAEL, another uncertainty factor (3 or 10) would be needed to adjust to a NOAEL.

As per the recommendations for the chronic table, a table of acute values should be developed, with the following columns created for each table: CAS, AEGL-1, etc. (as in the top row now, as modified based on the recommendations above). For each value, the year the value was last updated should be included and a qualitative description of the effect should be provided (e.g., describe the critical effect used as the basis of the calculation). Next, the table should be examined for consistencies. For example, if the

---

[16] Roach SA, Rappaport SM. But they are not thresholds: a critical analysis of the documentation of Threshold Limit Values. Am J Ind Med. 1990;17(6):727-53; Robinson JC, Paxman DG. The role of threshold limit values in U.S. air pollution policy. Am J Ind Med. 1992;21(3):383-96; Castleman BI. Legacy of corporate influence on threshold limit values and European response. Re: Am J Ind Med 44: 204-213, 2003. Am J Ind Med. 2006 Apr;49(4):307-9.

[17] Spacecraft Maximum Allowable Concentrations for Selected Airborne Contaminants: Volume 5 (2008) Committee on Spacecraft Exposure Guidelines, Committee on Toxicology, National Research Council (available at http://www.nap.edu/catalog.php?record_id=12529). All five volumes are available at www.nap.edu .

values from different agencies are different, the reasons should be explored. Perhaps one value is more recent than another; perhaps the critical effect is different. Such a table is complex and therefore a candidate for an appendix, with the summary result being in the main text.

Minor recommendations for clarification:

a. p. 2-13 bottom. The text should be revised to identify that the acute REL is for 1 hour. They also have 8 hour values, but we presume the analysis used the 1-hour values to make them equivalent to others.

b. p. 2-14 top description of AEGLs. In the middle of the paragraph, it says that the values range from 10 minutes to 8 hours. This is true, but they have explicit values for 10 minutes, 30 minutes, 1 hour, 4 hours, and 8 hours. Thus, the text should be expanded to indicate this. We presume the analysis used the 1 hour value for consistency.

The following minor edits are recommended. On page 2-16 Table 2-5:
(a) The table title should be revised to say 1-hour acute exposure.
(b) The table should be footnoted to define the AEGL-1, etc. (The definition is in the text, but tables should stand alone.)

Charge Question 4A

Section 2.2.3 describes the process by which we estimate chronic human inhalation exposures based on modeled average ambient concentrations at census block centroids. For these case studies, this process did not include consideration of either daily behavior pattern or long-term migration behavior. Section 2.2.3 presents a rationale for omitting daily behavior, and Appendix N presents a case study that adjusts inhalation-based lifetime cancer risk estimates for individuals to account for long-term migration.

4A Does our process of estimating inhalation exposures adequately support regulatory rulemaking? Is our rationale for omitting daily behavior convincing, or does the omission compromise the value of our assessments? Should this, or some other, adjustment for long-term migration be incorporated into our risk assessments?

Panel Response

An overarching concern with EPA's chronic inhalation exposure estimates is that children's exposures do not appear to have been adequately addressed. The differences in exposure between children and adults should be carefully considered and discussed in the exposure assessment. Otherwise, EPA's overall approach appears to be a reasonable *screening* approach for localized impacts (e.g. neglecting processes like deposition, plume depletion, atmospheric degradation) that can be refined further. In addition, EPA identifies some assumptions that could potentially lead to downward bias, such as not considering population growth or future expansion of production. Although these assumptions may be appropriate given the need to simplify the analysis, periodic reassessment may be needed, especially in circumstances where there are substantial changes in population growth and production levels.

With regard to the chronic inhalation exposure estimates, the Panel finds the rationale for omitting daily behavior to be convincing. Given the age of some of the available activity pattern data and the inherent community-scale activity pattern uncertainties between locations, the decision to omit daily behavior is justified. The Agency's draft RTR document should make it clear that consideration was given to daily behavior in terms of time spent indoors and past experience has shown it makes little difference in risk estimates.

The Panel further recommends that long-term migration not be incorporated into the risk assessment. It does not add value to the risk assessment and introduces additional uncertainty. As discussed in Appendix N, the migration data that would be used to

modify the risk estimates have not been scientifically peer-reviewed and are limited in their geographical representativeness. While this preliminary analysis does not merit being part of the central assessment, it is worth leaving in the appendix and referencing in the text.

Charge Question 4B

Appendix C describes a novel application of TRIM in the development of protective *de minimis* emission rates for 14 persistent and bioaccumulative HAPs (PB-HAPs). We believe that emissions below *de minimis* thresholds should not produce unacceptable risks in reasonable worst-case conditions. Facilities emitting below these values would not need to conduct a multi-pathway exposure and risk assessment.

Section 3.4 and Appendix I describe a refined application of the TRIM model in assessing multi-pathway pollutant transport and its subsequent impacts on human health from Portland cement facility air emissions identified as having a high potential to present significant impacts on human health.

We have limited our development of radionuclide risk estimates (described in Section 3.2.2 and Appendix G) to those associated with inhalation exposure. Radionuclides were not included in the multi-pathway risk assessment.

> 4B Is our use of the TRIM model to develop *de minimis* emission rates appropriate as a screening tool? Are the methodologies used in the refined multi-pathway assessment consistent with the best available science regarding multi-pathway pollutant transport and human exposures? Are the application of the model and the assumptions used clearly articulated? Are the resultant estimates of media concentrations and exposures clearly presented, explained, and interpreted? Given the large uncertainties surrounding the radionuclide inhalation assessment, are we justified in omitting radionuclides from the multi-pathway assessment?

Panel Response

Screening model framework and methodologies: In responding to this charge question, the Panel focused on how TRIM.Fate results were applied in the risk assessment process. The Panel did not evaluate the details of the equations in TRIM.Fate and did not itself evaluate the validity of the model. Appendix C describes a series of analyses that provide some confirmation that the screening model results are generally reasonable based on qualitative comparisons with environmental and food chain concentrations and partitioning, but these comparisons necessarily fall short of providing the level of confidence that could be gained by detailed comparison of model results and observations for a range of real-world applications. Appendix C indicates that EPA subsequently evaluated TRIM.FaTE's performance for modeling mercury and dioxins and furans, but does not discuss the results. As recommended by previous SAB panels,[18] we recommend that EPA continue to identify and acquire additional field data to estimate modeling parameters and to evaluate the TRIM.FaTE model components and other aspects of the modeling system on an ongoing basis. The NRC report *Models in Environmental Regulatory Decision Making*[19] provides useful guidance for these recommended efforts.

With the caution that continued efforts are needed to evaluate the TRIM.FaTE model, the Panel finds that EPA's screening approach is based on an appropriate framework and should provide a useful screen for sources that do not need a detailed site-specific multi-pathway analysis. The screening-level multi-pathway assessment is thorough and conservatively includes local subsistence agricultural and fishing scenarios, adding exposures across intake pathways to yield total PB-HAP exposure. Children and adults are modeled with doses calculated on an average daily dose (ADD) lifetime basis to assess chronic risk of these HAPs. This modeling is generally appropriate, although developmental and reproductive endpoints associated with mercury and dioxin can

[18] EPA-SAB-EC-ADV-99-003 (1998) Advisory on the Total Risk Integrated Methodology (TRIM), http://www.epa.gov/science1/pdf/eca9903.pdf; SAB-EC-ADV-00-004 (2000) Advisory on the Agency's "Total Risk Integrated Methodology (TRIM)," http://www.epa.gov/science1/pdf/ecadv04.pdf.
[19] *Models in Environmental Regulatory Decision Making.* Committee on Models in the Regulatory Decision Process, National Research Council, The National Academies Press, Washington, DC (267 pp, 2007).

involve shorter windows of vulnerability than lifetime exposure and so the dose rate averaging might need to be shorter for such endpoints. Use of an ADD may undervalue peak exposures that occur in early life or during pregnancy. Therefore, some discussion should be provided regarding whether consideration of early life windows of vulnerability and less than lifetime exposures should be considered.

Appendix C presents sensitivity analysis results to identify the most influential input variables in the screening assessment. As EPA recognizes, facilities with emissions exceeding the screening level thresholds might end up doing so because of assumptions in one particular area (e.g., soil to vegetation uptake rate; beef biotransfer factors; fish ingestion). This analysis could be refined so that these particular factors are evaluated in a distributional sense to enable Monte Carlo analysis, leading to an overall multi-pathway probability distribution of risk rather than a bright line estimate. In this way, the probability that a facility's emissions could lead to unacceptable risk could be estimated and presented to risk managers to weigh against other factors.

Communication of assumptions and results: The Panel considered it reasonable for the Agency to set an emission threshold below which detailed site-specific multi-pathway analysis (including potentially extensive data collection) would not be necessary for each source. However, the choice of the term "*de minimis*" to describe this threshold was unfortunate, as it obscures the conclusions of the near-source multi-pathway analysis. In particular, when the *background* concentration of a PB-HAP already exceeds a safe level (e.g., where a fish advisory is already in effect) the public may not understand a local source's contribution being characterized as *de minimis*. Additionally, although such risk may not be deemed "unacceptable", it is not clear that a threshold set at a 1 in 1 million cancer risk or chronic HQ of 1 should be characterized as *de minimis* in the presence of elevated background contributions.

Risk assessments must be credible to the public. Exhibit 4-7 showed modeled concentrations in sediment and surface water for the screening scenario that were higher than most of the values from the literature. For example, in the screening scenario the modeled concentration in sediment is about an order of magnitude higher than reported for Minnesota lakes, and Minnesota has a statewide fish advisory for Hg. Thus EPA's

finding that the corresponding Hg^{+2} emissions rate of 1.6E-01 TPY (320 lbs) (Exhibit 2-3) is "below a level of concern" may not be credible to the public.

Instead of "de minimis emissions levels", it would be better to describe EPA's screening model results as providing an "action threshold for local hot-spot analysis." Using a model to estimate the relative contributions of local and background sources of a pollutant is useful for informing policy choices and communicating with the public. However, the model results need to be clearly presented to show 1) the relative fraction of the local source's emissions that are deposited locally versus being transported to add to regional burdens, and 2) the relative contributions to total multi-pathway exposure from local and regional background sources. If the local source contribution is small relative to background, refined site-specific modeling would provide little information beyond what could be obtained from a regional or national-scale analysis, so screening out individual sources from further analysis is appropriate. Nevertheless, the contribution the source category makes to overall emissions of PB-HAPs should still be considered. From a scientific standpoint, EPA must also ensure that ignoring background pollutant levels of PB-HAPs does not lead to incorrect results due to nonlinear physical and chemical processes in the fate and transport model. Where nonlinear processes are at issue, individual source contributions can be tagged for tracking, but all contributions including "background" must be considered in the fate and transport model.

Previous SAB review panels have similarly recommended that EPA characterize background as well as incremental risks in its residual risk assessments. Quoting from the SAB Advisory on the USEPA's Draft Case Study Analysis of the Residual Risk of Secondary Lead Smelters (p. 11) "[a] residual risk analysis that does not add exposures to baseline contamination to the estimates of on-going contamination may vastly underestimate the hazard quotient at the site and incorrectly conclude that the on-going releases pose risks at less than threshold levels."[20] The Secondary Lead Smelters review

[20] See EPA-SAB-EC-ADV-00-005 (2000) An SAB Advisory On The US EPA's Draft Case Study Analysis Of The Residual Risk Of Secondary Lead Smelters

also noted (p. 25) "The [Residual Risk] Report to Congress (USEPA, 1999)[21] discusses the need to include background risk and the difficulty associated with this specific issue. … The absence of an assessment of background risk seriously impacts statements about the conservative nature of the refined screening assessment." Our Panel concurs with these comments. The need to characterize background as well as incremental risks also arises in the case of some non-PB-HAPs such as benzene, but the issue stands out for the PB-HAPs because of their nature as persistent and bioaccumulative and because for most pollutants evaluated with EPA's screening scenario, a large fraction of the emitted mass was lost from the model domain through advection downwind (See Exhibit 4-1, Appendix C).

Omission of radionuclides from the multi-pathway assessment: Local impacts of radionuclides, including naturally occurring isotopes, need to be considered based on better data for radionuclide concentrations in geological feed materials to mineral processing industries. The comprehensive analysis presented in Leenhouts (1996) and the results of the Portland cement case study suggest that radionuclide emissions may be a risk for any industry category that grinds and heats large amounts of natural mineral feedstock. Radionuclides need to be considered in the residual risk assessment process, but as discussed in response to charge question 1C, preliminary work is needed before attempting to use TRIM. There is currently no reporting of actual radioactive isotope type and unit of radioactivity for Portland cement feedstocks.

At this early stage of the assessment of radionuclide emissions, the Panel agrees it is acceptable to omit the multi-media assessment. Ultimately, however, a multi-pathway assesment is needed because non-inhalation pathways are often important for radionuclides that can accumulate in biota and subsequently be ingested. Radon is not likely to bioaccumulate as it is an inert gas, but the fate of its decay products need to be considered. The literature on ^{210}Po in the food chain needs to be reviewed to determine if it bioaccumulates. The literature on multi-pathway exposure from ^{210}Po in phosphate fertilizer may provide information on this issue.

[21] See EPA-453/RR-99-001 (1999) US EPA's Report to Congress on Residual Risk

Particle bound HAPs: A potentially serious omission from the Appendix C analysis is the issue of HAPs associated with coarse fraction (PM2.5-10μm) and very coarse (> 10μm) particles. Large particles deposit rapidly, thus causing relatively high impacts near a source. If the HAPs-containing particles are injected into the air near ground level (fugitive emissions and resuspended road dust) then the fraction deposited nearby is much higher compared to the same particles being emitted from a stack. The methodologies used in the case studies would not detect local multi-pathway risk caused by deposition of particle-bound HAPs near the source site.

Charge Question 5

Section 2.2.5 describes our process for developing screening and refined estimates of acute inhalation risk. For acute screening purposes we have assumed that, in the worst case, a person could be exposed for one hour to ten times the highest hourly concentration calculated by the dispersion model. This in effect assumes a 1-hour emission rate of ten times the annual average (assuming continuous emissions), simultaneous occurrence of "worst-case" meteorological conditions, and also the presence of a person at this worst-case downwind location.

Appendix B presents an effort to evaluate the protectiveness of this screening assumption using detailed short-term emission data for a limited geographic area. Appendix E describes our refinement of acute risk estimates for refineries that failed the acute 10X screen, by using more accurate emission points and property boundaries.

Our refined acute assessments do not combine acute hazard quotients associated with different HAPs because of the inconsistent nature of acute health benchmarks and the inherent conservatism of our exposure assumptions.

5 Does the 10X acute screening assumption for petroleum refineries appear to be appropriately protective? If not, is it under- or over-protective? Given that this analysis applies only to sources in the Houston area, can we apply the 10X assumption to HAPs in

other source categories or should we consider some other approach for some other HAPs, *e.g.*, metals? Is there some other way we might address high emission events such as startup or shutdown of processes? Are the refinements to the acute screening assessment objectively employed and scientifically defensible? Should we sum acute hazard quotients by target organ in the same way we do for chronic hazard quotients, *i.e.*, a target organ specific hazard index (TOSHI) approach, or are our reasons for not doing so adequate?

Panel Response

Use of 10X annual emissions for short-term exposure estimates: The Panel agreed there is a critical need for better data addressing short-term exposures to HAPs. However, in the absence of chemical- and site-specific data, the use of the 10X screening assumption for petroleum refineries seems reasonable, taking into consideration the assumption of simultaneous occurrence of 'worst-case' meteorological conditions and the presence of a person at this worst-case downwind location.

While the Panel supports the use of the 10X assumption in the absence of better information, the methods used to derive and justify the 10X screening assumption are not readily apparent from Appendix B. The authors should consider using a more transparent approach to presenting this data. In revising Appendix B, EPA should at least explain more clearly why the median and mean values of event to long-term release rates are less than 1. Furthermore, the figures contained in the referenced reports by Allen et al. provide an easily understandable template that could be used to present the development of the 10X screening assumption used to assess acute impacts.[22,23] Figures 2 through 8 of the Allen et al. paper clearly show the baseline annual hourly emission rates for VOCs,

[22] Allen D, Murphy C, Kimura Y ,Vizure W Jeffries H, Kim, B Webster M and Symons, M. Variable Industrial VOC Emissions and their Impact on Ozone Formation in the Houston Galveston Area. Available at http://www.epa.gov/ttn/chief/conference/ei13/uncertainty/allen.pdf

[23] Allen D Murphy C, Kimura Y ,Vizure W Jeffries H, Kim, B Webster M and Symons, M. Variable Industrial VOC Emissions and their Impact on Ozone Formation in the Houston Galveston Area. (April 2004). Final Report Texas Environmental Research Consortium Project H-13. Available at http://files.harc.edu/Projects/AirQuality/Projects/H013.2003/H13FinalReport.pdf

highly reactive VOCs and 1,3- butadiene, and the magnitudes of the excursions over the baseline annual hourly emission rates. The results demonstrate that the facilities in the Houston-Galveston area clearly do not have emissions that are constant and continuous. The daily emissions can vary from the annual average emissions by a factor of 10 to 1000. Figure 1 of the Allen et al. paper also provides a useful conceptual illustration of the four characteristic types of emissions variability from the industrial sources in the Houston-Galveston area which EPA might adapt.[18]

Apart from our concerns about data presentation, the Panel concurred that a release factor associated with the 99th percentile value would seem to be appropriately health-protective. However, there is one significant limitation to the TCEQ database that needs to be identified in Appendix B, which is that the emissions event reporting rule only requires reporting from the time of discovery until the event was corrected. This would cause a low bias for both the event duration and quantity of emissions released.

In Appendix B, EPA attempts to address the representativeness of the Texas Commission on Environmental Quality (TCEQ) data base by filtering the data to isolate routine and allowable hazardous air pollutant (HAP) excursions from major emitters in the Houston-Galveston area (Table 2 of Appendix B). There appears to be a mixture of source types (e.g., petroleum refineries and chemical manufacturing plants) in Table 2; isolation of petroleum refinery specific allowable hourly excursion data in the manner described above would provide a more transparent justification of the conservative nature of the 10X screening assumption. We are also concerned about the filter applied to attempt to remove facilities below the major threshold from this analysis. NESHAP applicability for each identified facility should be easily obtained from the current Title V permits, the EPA should revisit this filtering assumption to insure all facilities subject to the NESHAP are included in the analysis. The Panel also suggests that following the screening process, the chemicals of highest concern (drivers) be evaluated against the list of chemicals reported in the Houston area (Appendix B), to ensure they are adequately represented.

The 10X screening assumption makes the most sense under conditions when the production facility is operating continuously (24 hrs/day and 7 days/week) for the entire year. However, adjustments may need to be made for other source categories where facilities operate during only part of the day or part of the year. Under these scenarios, which may frequently occur as demonstrated by the Allen et al. report and papers, estimates of daily releases calculated from annual release values may seriously underestimate releases occurring during production periods.[18, 19]

Although the Panel generally agreed that the 10X assumption could be used for other geographic areas, it was felt that the actual releases would be dependent upon the manufacturing processes involved which may or may not be captured in the Houston example. There was limited information on the manufacturing processes in the Houston area included in the document making it difficult to evaluate its relevance to the case studies. We would recommend that a table listing the industries by Standard Industrial Codes (SIC) be included in the evaluation to allow comparisons with industries to be evaluated in the future. The report by Allen et al. (2004) reports these excursions by SIC codes, allowing for some understanding of source category-specific emissions variability.[2] In the case of petroleum refineries, for example, it appears there are four types of emissions points (fugitives, pipelines, cooling towers and flares) associated with the short-term excursions.

Going beyond the Houston-Galveston data set, the Panel suggests that an estimate of the variation and peaks in short-term emission rates could be obtained by examining time trend data from continuous emission monitors. Since HAPs emitted from a stack are often controlled by the same air pollution equipment that is used for criteria pollutants, it may be reasonable to use variations in PM, SO_2, or opacity as surrogates for variation in emission of HAPs. Another option may be the utilization of real-time fenceline measurements (FTIR or UV-DOAS) collected during enforcement and research investigations around facilities such as petroleum refineries.

Refinements to acute screening assessment: For facilities with elevated acute hazard quotients, EPA used aerial photographs to ensure exposures were not occurring within property boundaries, and re-estimated maximum (off-site) exposures where this had occurred. The Panel found this to be a useful refinement. In addition, however, where pollutants emerge as drivers of acute risks, EPA should also re-examine the acute toxicity reference values used in the assessment to make sure that they are correct and appropriate for the assumed 1-hr period of exposure. For example, the acute REL for benzene used in the case studies (1.3 mg/m^3) appears to be based upon a 6-hour exposure period rather than a 1-hr exposure.[24]

Summing acute hazard quotients: The Panel recommends that EPA examine the likelihood that a 10X release would occur under the most hazardous meteorological conditions and how likely it would be for 10X releases of multiple chemicals to occur simultaneously. If it is concluded that simultaneous releases under adverse meteorological conditions would be very unlikely, then summing the acute hazard quotients by target organ would not be necessary. Alternatively, screening could be done using the TOSHI approach with more detailed follow-up for agents or combinations of agents that were identified as potential concerns. A primary focus of this approach would be on irritants, which are generally of most concern for acute exposures.

Charge Question 6

Section 3.5 and Appendix J describe a refined, site-specific application of TRIM to conduct an ecological risk assessment for PB-HAPs emitted by the same Portland cement facility evaluated in the human health risk assessment. Appendix J also describes a nationwide facility ranking exercise that identifies Portland cement facilities with the highest potential for causing indirect ecological effects via acidification of the environment by hydrogen chloride emissions. Appendix K describes an analysis of

[24] OEHHA (1999) http://www.*oehha.ca.gov/air/acute_rels/pdf/71432A.pdf*
Note: This information can also be found in summary form in the reference (http://www.oehha.ca.gov/air/pdf/acuterel.pdf) cited in the text on page 2-13.

possible direct effects on plant foliage of air concentrations of hydrogen chloride emitted from Portland cement facilities that are below human health thresholds.

6. Is the ecological assessment case study scientifically defensible? Does it conform to EPA risk assessment guidance (*e.g.*, *Guidelines for Ecological Risk Assessment*, *Risk Characterization Handbook*, *etc.*)? If not, how can we improve it? Are the elements of the ranking scheme adequate to identify the facilities most likely to be of concern? Are there better data sources or approaches for drawing conclusions for specific locations? With regard to investigating the potential for direct ecological effects at air concentrations below human health thresholds from other sources or source categories, what suggestions can be made for prioritizing additional HAPs for literature searches similar to that done for hydrogen chloride in Appendix K?

Panel Response

Ecological risk assessment case study: The ecological risk assessment (ERA) presented in Appendix J is an impressive effort tackling an extremely complex issue. While it is a good first step, the ERA needs to be improved, as the EPA ERA guidelines were not followed well. This would entail doing a problem formulation stage (which is perhaps the most important stage; see Dale et al. 2008[25]) that shows the ecological conceptual model, which then directs the study design, and subsequent linkage between assessment and measurement endpoints. In addition, the risk characterization did not show how the measurement endpoints linked back to the assessment endpoints and conceptual model. Each stage of the ERA (problem formulation, exposure/effects characterization, and risk characterization) can be improved, with specific suggestions given below.

[25] Dale, VH et al. 2008. Enhancing the Ecological Risk Assessment Process. Integrated Environmental Assessment and Management 4:306-313. [SAB report entitled, "Advice to EPA on Advancing the Science and Application of Ecological Risk Assessment in Environmental Decision Making: A Report of the U.S. EPA Science Advisory Board", available at: http://yosemite.epa.gov/sab/sabproduct.nsf/7140DC0E56EB148A8525737900043063/$File/sab-08-002.pdf]

The selection of contaminants of concern for the ERA case study (Appendix J, section 2.1) is well justified as is the choice of key ecological receptors (Appendix J, section 2.3.1). However, given the paucity of information on other potential HAPs, a separate research effort is warranted to rank HAPs for analysis. This effort should consider particulate-associated HAPs, high K_{ow} compounds, and multiple exposure pathways as shown on the flow charts EPA presented to the Panel to summarize its approach for RTR health assessments.

The heavy reliance of the ERA case study on TRIM.FaTE is a concern, as this EPA model has not been well validated in the peer-reviewed literature for ERAs, and an adequate sensitivity analysis with ground-truthing is lacking. A related concern is the fact that all exposure and effects predictions are based on generalized assumptions, and as discussed in response to Charge Question 1A, there are multiple indications that emissions may be underestimated. The potential for error propagation is a concern. More transparency is needed for key parameters used in TRIM.FaTE for the ecological (as opposed to the human health) risk assessment, such as sediment concentrations of the chemicals of potential concern and whether or not (and how) their bioavailability is linked to key factors (e.g., total organic carbon (TOC), dissolved organic carbon (DOC), and hardness). Appendix I (referenced as the source of the information) is confusing in this regard. It appears virtually all TRIM.FaTE parameters for the test site have been estimated and extrapolated from other sites with a significant amount of "professional judgment", making this a truly hypothetical ERA. This raises the question of how can we assume there is no risk for this, much less other Portland cement facilities, without some degree of verification that the model's predictions regarding food web, chemical fate and speciation, biological uptake and effects are correct?

From the information presented in Appendix J, the case study appears to have relied on Toxicity Reference Values (TRVs) based on data that are 15 years old or older. In addition, it is difficult to determine if, or which, data from Appendix J were used. There have been a multitude of excellent peer-reviewed studies that are relevant to this process, as they have focused on mercury and highly chlorinated compounds such as dioxins.

(Relevant examples are listed at the end of this response.) There are recent data for the Housatonic and Hudson Rivers for TCDD congeners (and related PCBs), that could be further employed to reinforce the assumed concentrations and feeding patterns.

In Appendix J, section 3.2.3, EPA discusses and rejects the option of using TRVs expressed in terms of tissue concentrations instead of chemical intake. However, reporting TRVs in terms of tissue concentrations (rather than intake as commonly done for human risk assessments) would allow for more and better comparisons with the peer-reviewed literature and predictions of risk, as there are fewer peer-reviewed literature reports of intake values. The Agency's draft RTR document should add tables of these values and calculate new HQs based on steady-state tissue concentrations. It would also be helpful to see predicted concentrations in sediment and sediment quality guidelines listed in the same table.

In the case study, EPA uses a two-stage approach to characterize ecological risks from Hg and dioxin emissions. In the first stage, risks are summarized by computing hazard quotients as the exposure dose divided by the TRV (Appendix J, section 3.2.4). The Hazard Quotient approach is justifiable as a crude screening level approach in applications such as the RTR assessments, but only if very conservative values are utilized. For ground-truthing this effort, or for a refined ERA, probabilistic approaches are needed. We know Hazard Quotient-type ERAs are fraught with unacceptably high levels of uncertainty regarding exposures and their linkages to adverse effects and do not account for multi-stressor and non-chemical stressor interactions and resulting effects. The assumptions of constant exposures are of course conservative, but best used in a "reference condition" approach whereby multiple reference sites within the area of the facility are considered. This will help account for the non-facility related exposures and effects. For a more thorough discussion of these issues and others important to improving the ERA process, see Dale et al's (2008) summary of their recent EPA SAB report (EPA-SAB-08-002).[18]

For ecological risk, there are some overlapping charge questions with the human health risk assessment that should be considered. In particular, Charge Questions 2 and 3, concerning dispersion modeling and dose-response assessment, affect both risk assessments. The environmental chemistry (atmospheric chemistry) and fate are critical for ecological assessment endpoints to be determined. For example, more consideration needs to be given to how particulate matter may interact with certain types of chemicals in the emissions. In particular, coarse particulates that settle within 1 to 2 miles of the site may contain high levels of contaminants and should be considered as a potential exposure compartment. High K_{ow} compounds, such as PAHs and dioxins, will adsorb to carbon, so the presence of particulate matter may be critical in bioavailability and fate. QSAR (Quantitative structure and activity relationship) models are important in this respect, as is the nature of the ecosystem into which the chemicals and particulate matter deposit. Although the TRIM-FaTE model simulations indicate little expected risk to humans via inhalation, other receptor organisms, such as benthic macroinvertebrates and fish in waters or soil invertebrates may be affected.

On pages J-29 through J-33, EPA presents a sensitivity analysis of how angler harvesting would affect mercury and dioxin concentrations in food web compartments for the ERA case study. This analysis should be omitted, as fish harvesting by fishermen should not be a part of an ERA.

In characterizing the risk for the Ravena case study (page J-46), EPA discounts the finding of elevated HQ values for Ravena pond on the grounds that it is a small water body with correspondingly small wildlife populations. The rationale that no population effects will occur in a small water body because there are few individuals is flawed. In fact, small water bodies with small populations may pose special concerns, as has been shown in prior studies. Greater population effects would be expected in systems with fewer individuals, particularly with limited to no refugia for recruitment. Page J-46 also indicates no adverse effects are expected for piscivorous and insectivorous wildlife, even though they have elevated HQs. The stated conclusions cannot be justified using the HQ approach.

The assumption that ecological receptors will be protected if human health is protected is incorrect. Recall the "canary in the coal mine" approach was derived long ago and it is well known that wildlife are good sentinels for protecting humans due to their greater sensitivity. Through literature comparisons it should be possible to develop a sound "safety/application" factor that protects species of concern (note mink/otter will likely be the species of greatest risk). This literature based factor could then be used to back calculate (via TRIM-FaTE) to an allowable emissions concentration, which would fit nicely into the existing flowchart replacing the top decision point based on no human health effects.

Facility ranking scheme: The process to select the Portland cement facilities of greatest potential concern for HCl deposition using pH, hardness, alkalinity and soil type data was very good. For site-specific ERAs, however, other site characteristics should also be considered, such as altitude, gradient, trophic status, TOC levels, watershed location (e.g., headwaters), sensitive land uses (forested, protected areas, wetlands), and sensitive, threatened or endangered receptors (e.g., amphibians, mussels, piscivorous wildlife). For the discussions on Hg, the trophic status of the receiving lake or pond becomes important. Methylation of Hg is very site dependent. For example, it tends to be stronger in lakes with high organic matter in the sediments. Over time, much of the terrestrial primary production moves to the aquatic habitats in watersheds. Hence, the buildup of organic materials in shallow riparian habitats influences the bioavailability of chemicals deposited. If the RTR process is to establish a guide for ERAs conducted under the Clean Air Act, there may be value in adding a section on the importance of obtaining, for each site, site-specific emissions and exposure data. Otherwise, it will be difficult to account for the wide range of critical factors that will affect ecological risk and are defensible in a court of law.

Direct-contact ecological assessments: EPA explains it has not developed criteria for HAPS for direct-contact ecological assessments (page 3-20), yet there was an RTP workshop 3 to 5 years ago [Federal Register Notice announcement published

September 8, 2005 (Volume 70, Number 173,Page 53360)] with the focus of bringing the ERA process into emissions of HAPs. There were a lot of good ideas put forward that should be considered for the RTR assessments. There should be a peer-reviewed effort to reevaluate other potential HAPs of ecological concern, particularly those that associated with particulates, from both petroleum refinery and Portland cement operation emissions.

In summary, many of the above concerns and issues can be addressed by conducting a ground-truthing ERA at a site such as the Ravena Pond, or by a comparison of TRIM.FaTE, predictions with more conventional ERA methods (e.g., using Bioaccumulation Sediment Accumulation Factors in food web models (e.g., TrophicTrace and EcoFRAM by Frank Gobas, USEPA's AQUATOX 2.2, CATs) at a well studied site with similar CoCs (e.g., see web sites for USEPA reports on Superfund sites: Lower Housatonic River, Region I (http://www.epa.gov/NE/ge/pcbshealthandenviro.html); Hudson River, Region II (http://www.epa.gov/hudson/reports.htm); Fox River, Region V (http://www.epa.gov/Region5/sites/foxriver/index.html)). This could be done by a contractor experienced with ERAs and they could use more conventional fate/effect ERA models using both deterministic and probabilistic approaches with limited on-site sampling of exposure compartments and receptors. Sediment concentrations of the CoCs can be linked to food web bioaccumulation and then compared to adverse tissue levels in the key receptors. This rather simple effort would determine whether the proposed generalized approach works and is of sufficient accuracy to warrant its nationwide application. This would allow for refinement of the "nationwide" Tier 1 ERA approach and with general guidelines for site-specific, Tier 2-type evaluations.

Some references that may be relevant as EPA reviews its approach for ERA in the RTR process are provided below.

Bargar TA, Scott GI and Cobb GP. 2001. Maternal transfer of contaminants: Case study of the excretion of three polychlorinated biphenyl congeners and technical grade endosulfan into eggs by white leghorn chickens (*Gallus domesticus*). *Environmental Toxicology and Chemistry* 20:61-67.

Brasson RL and Cristol DA. 2008. Effects of mercury exposure on the reproductive success of tree swallows (*Tachycineta bicolor*). *Ecotoxicology* 17:133-141

Custer TW and Heinz GH. 1980. Reproductive success and nest attentiveness of mallard ducks fed Aroclor 1254. *Environmental Pollution* (Series A) 21:313-318.

Custer TW et al. 2002. Dioxins and congener-specific polychlorinated biphenyls in three avian species from the Wisconsin River, Wisconsin. *Environmental Pollution*. 119:323-332.

Fernie KJ, Smits JE, Bortolotti GR, and Bird DM. 2001. Reproductive success of American kestrels exposed to dietary polychlorinated biphenyls. *Environmental Toxicology and Chemistry* 20:776-781.

Hochstein JR, Bursian SJ, Aulerich RJ. 1998. Effects of dietary exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin in adult female mink (Mustela vison), *Archives of Environmental Contamination and Toxicology*, 35(2), 348-53.

Ludwig JP, Kurita-Matsuba H, Auman HJ, Ludwig ME, Summer CL, Giesy JP, Tillitt DE and Jones PD. 2009. Deformities, PCBs, and TCDD-Equivalents in Double-Crested Cormorants (*Phalacrocorax auritus*) and Caspian Terns (*Hydroprogne caspia*) of the Upper Great Lakes 1986–1991: Testing a Cause-Effect Hypothesis. Available online 23 February 2009.

McCarty JP and Secord AL. 1999. Reproductive ecology of tree swallows (*Tachycineta bicolor*) with high levels of polychlorinated biphenyl contamination. *Environmental Toxicology and Chemistry* 18:1433-1439.

Nosek JA, Craven SR, Sullivan JR, Olson JR and Peterson RE. 1992. Metabolism and disposition of 2,3,7,8-tetrachlorodibenzo-p-dioxin in ringnecked pheasant hens, chicks, and eggs. *Journal of Toxicology and Environmental Health* 35:153-164.

Rice CP et al. 2003. Souces, Pathways, and Effects of PCBs, Dioxins, and Dibenzofurans. In, Hoffman DJ et al. (eds), Handbook of Ecotoxicology, 2nd edition. CRC Press. Boca Raton FL. Pp 501-573. (a critical review article)

Tanabe S, Subramanian A, Hidaka H and Tatsukawa R. 1986. Transfer rates and pattern of PCB isomers and congeners and pp-DDE from mother to egg in Adelie penguin (*Pygoscelis adeliae*). *Chemosphere* 15:343-351.

Tillitt DE, Gale RW, Meadows JC, Zajicek JL, Peterman PH, Heaton SN, Jones PD, Bursian SJ, Kubiak TJ, Giesy JP, and Aulerich RJ. 1995. Dietary Exposure of Mink to Carp from Saginaw Bay. 3. Characterization of Dietary Exposure to Planar Halogenated Hydrocarbons, Dioxin Equivalents, and Biomagnification. *Environ. Sci. Technol.*, 30 (1), pp 283–291.(Publication Date (Web): December 27, 1995).

Wiener JG et al. 2003. Ecotoxicology of Mercury. In, Hoffman DJ et al. (eds), Handbook of Ecotoxicology, 2nd edition. CRC Press. Boca Raton FL. Pp 409-463. (a critical review article)

Charge Question 7

The risk characterizations for these two case studies (Sections 2.3 and 3.6) represent our current practices in providing information to decision-makers responsible for RTR rulemaking. The analyses presented in the appendices are by and large illustrative of what can currently be done in the regulatory context, given time and resource constraints.

7 Do these characterizations objectively and completely incorporate the goals and principles of EPA's *Risk Characterization Handbook* to the extent scientifically feasible? In particular do they provide a complete and transparent discussion of uncertainties and limitations? If not, how can the risk characterizations be improved? Can you suggest where we might focus any additional efforts and resources in order to have the biggest

impact on refining risk characterizations for these RTR assessments, ultimately leading to better regulatory decision-making?

Panel Response

Risk characterizations are often difficult to develop because a highly technical assessment must be communicated to decision makers and others who may lack some of the underlying technical background. As stated in the *Risk Characterization Handbook* (pg 13):

> **"Are Risk Assessment and Risk Characterization the Same?**
> No, they're not the same. Risk assessment is a process comprised of several steps (see section 1.2.1 above for detail). Risk characterization is the culminating step of the risk assessment process. Risk characterization communicates the key findings and the strengths and weaknesses of the assessment through a conscious and deliberate transparent effort to bring all the important considerations about risk into an integrated analysis by being clear, consistent and reasonable".

The Agency's draft RTR document took great care in summarizing and providing justification/explanation for most of the results, including attention to uncertainties. The summary tables (tables 2-7 and 3-3) were well done and provide a concise summary of the risk assessment results for the risk manager. However, a number of improvements are possible.

Presentation of risk characterizations: In the RTR case studies, the presentation of methods, risk assessment results, and risk characterization are intermingled, such that the purposes of the risk characterization are not met. This can be improved by focusing more on the purpose of the characterization to communicate with decision makers as the primary audience, recognizing that transparency is important and that the audience will inevitably be broad (e.g., a reporter may use it as a source for a story, the regulated

source may use it for community interaction). To these ends the Panel recommends the following improvements:

1. Develop a separate methods document that contains a full description (including uncertainties) of all of the common components of the source-specific risk assessments. For example, it would include EPA RfC and cancer assessment methodologies, the National Emissions Inventory (NEI) description, AEGL methods, etc.

2. Refer back to this master document, as appropriate, in source-specific risk characterizations, while providing additional information particular to the source category at issue. For example, in a source-specific risk characterization, there is no need to repeat a discussion of mode-of-action for cancer risk if it wasn't used. On the other hand, source-specific discussions of uncertainties are far more useful than generic boilerplate about uncertainties. For example, there may be particularly strong (or weak) elements of the emissions inventory that need to be discussed for a specific source.

3. While other sections of the RTR assessments should document the technical details, the risk characterization sections should stand alone. A broad outline of the risk characterization section would include:
    a. The general background information for the RTR assessment (perhaps using flow diagrams).
    b. The risk characterization, with sections on emissions, cancer risk, non-cancer risk, and ecological risk, each of which *integrate* results and uncertainties and are written for EPA decision-makers. For HAPs that are found to drive risks, the risk characterization should include expanded discussion of the nature of the effects at issue (including qualitative cancer classification if applicable) and potential susceptibilities (e.g., children, elderly, women of

childbearing age, individuals with preexisting diseases). For example, page 3-23 says that the maximum hazard index for Portland cement manufacturing is associated with potential effects of manganese compounds on the central nervous system. But what type of CNS effects are they? What groups are expected to be more susceptible? Expanded discussion is important to understanding the "real-world" risk, including dealing with health disparities. For example, it would be important to recognize if a risk driver for a particular facility exacerbated asthma and the community surrounding the facility was a low-income population with an elevated incidence of asthma.

c. A summary with a clear description of risks of concern, using language understandable to an educated non-technical audience. This section should be relatively brief and balance the weight-of-the-evidence. An example audience for this section might include officials in the community where a facility of concern is located; they should be able to understand the nature of and degree of risk to human and ecological health.

Risk characterization for facilities covered by more than one source category: The Clean Air Act requires residual risk assessment of source categories, which have a particular definition that may only include part of a facility. For example, the petroleum refinery MACT 1 standards do not cover combustion processes within a refinery facility. Although this requirement for separate assessments has practical regulatory implications, it only partially accounts for potential human health or ecological risk. Since regulators seek protection of the public health and the environment, the risk characterization should clearly explain the inherent limitations of only dealing with one source category at a time. This limitation needs to be clearly noted for the risk manager. This will not change the source risk characterization itself. However, it can change its interpretation, especially in the case of large industrial complexes. For example, the Coke Oven Residual Risk Assessment clearly identifies that it is assessing a source

category (i.e., coke batteries) that is only part of an entire facility.[26] The risk assessment provides the estimated risk associated with emissions from the subpart and also accounts for similar emissions from different processes at the plant to provide the risk manager with an estimate of the total facility risk in the surrounding community. The risk characterization section should provide an estimate of total facility risk for facilities subject to multiple federal emission standards for hazardous air pollutants or clearly identify it as an outstanding issue that needs to be examined further.

Characterization of aggregate and cumulative risks: Since risk will be influenced by aggregate and cumulative exposures, finding that a source category has no significant risk from a particular chemical or a mixture of chemicals does not mean that people in the area are without risk from that chemical or mixture. For example, Houston faces particularly difficult air toxics challenges due to the significant air emissions from one of the largest petrochemical complexes in the world. There are more than 100 benzene sources alone from refineries and chemical plants in the Houston area. Harris County, in which Houston is located, over 19 million pounds of hazardous air pollutants were emitted in 2003, including 750,325 pounds of benzene according to the EPA's 2003 Toxic Release Inventory (TRI) report. From a public health viewpoint, personal exposures resulting from occupational or behavior (e.g., smoking) sources can also contribute to risk beyond that of a particular source category.

Ecological examples where aggregate risks are important also exist. The PB-HAP methodology used for the case studies does not consider background concentrations, focusing instead on incremental risks from the source category. However, ecological resources are also influenced by aggregate and cumulative exposure that must be considered in protecting the environment. The ecological receptors, just as humans, are affected by their total environment and all the stressors to which they are exposed. Exposures to multiple stressors at sub-lethal levels can result in lethal effects. Since many of these facilities will be located in human-dominated watersheds, there is a high probability that organisms will be exposed to multiple stressors.

[26] United States Environmental Protection Agency (USEPA) 2005a. USEPA Risk Assessment Document for Coke Oven MACT (maximum achievable control technology) Residual Risk – March 31, 2005

Linkage of Hazardous Air Pollution Emissions and Risk Assessment: RTR assessments must provide clear documentation of the hazardous air pollutants emissions that are modeled in the risk assessment. For example, the RTR case study models actual emissions using the 2002 National Emissions Inventory (NEI) and there apparently was an adjustment of these emissions using site-specific data from 22 refineries as provided by the American Petroleum Institute. However, it is not clear what adjustments were made.

The risk characterization for petroleum refineries includes a discussion of an ingestion pathway screen for POM emissions and indicates all 156 facilities were screened. However, one important aspect of this screening was not adequately explained. Only 70 facilities reported some type of POM emissions (Table 2-6) and the POM emission rates used to assess the potential risk for 156 facilities are never explained to the risk manager. The emission summary tables should include the emissions that were modeled to estimate cancer and non-cancer risk for the inhalation and ingestion pathways.

Identification and discussion of uncertainties: RTR assessments must proceed even though most will have a relatively long list of uncertainties. Such a list should be treated as an opportunity to identify future improvements. Insofar as possible, the Panel recommends that EPA perform a sensitivity analysis to identify the major uncertainties and then proceed to: (1) explain them clearly in the risk characterization section and (2) take steps to reduce them. For example, it appears that the NEI and paucity of up-to-date IRIS values are very likely to have a significant impact on the quality of the RTR assessments. These problems should be emphasized more, and EPA management should seek improvements so that future assessments can benefit. Action on major uncertainties that can be identified very early in the assessment of a particular source could have a substantial impact on the utility of that assessment. For example, if an apparent chemical driver has an out-of-date (or no) health value, it may be possible to rectify this problem prior to completion of the assessment.

In the case studies, EPA has generally done a good job of investigating uncertainty in many aspects of the inputs for the residual risk assessments. The sensitivity analyses provided to the Panel are extensive. However, the next step is to carry the results of these uncertainty analyses through to the final risk results and characterization. There are too many sources of uncertainty to qualitatively brush aside differences with statements discounting degrees of uncertainty because of either the risk range or that the component in question "does not introduce significant uncertainty into the risk assessment relative to other sources of uncertainty that limit reporting risk estimates to one significant figure" (page 4-7).

As one example, on pages 2-21 and 2-27, the concern is raised that Canadian and European studies [30, 31] suggest emissions from some refineries are significantly higher than amounts estimated, but this is acknowledged almost in passing in the uncertainty section (page 2-30). This is an important issue for the human and ecological risk assessment and if it cannot be accounted for, then appropriate uncertainty factors must be used.

Ecological risk characterization: According to USEPA ERA guidelines, the Risk Characterization should link measurement endpoint effects back to the assessment endpoints and conceptual model, which the Agency's draft RTR document does not do. Furthermore, as discussed in the response to Charge Question 6, no site-specific data exist for the ERAs, with every aspect of the ERA from exposure to effects (thus the risk characterization) being based on non-site data averaging, assumptions, questionable extrapolations, averaged/steady state conditions, and literature-based values. It seems that the only ways to get past the huge uncertainties involved are to do some case studies with site specific data that would represent a Tier 1 ERA, or to compare the TRIM.FaTE predictions to those of another peer-reviewed study (e.g., studies conducted at Superfund sites such as the Lower Housatonic River, Fox River, or Hudson River). After those efforts are completed, a guidance document could be provided that explains the general ERA process for the Tier 1 exercise, the most critical input parameters to determine if a site risk may exist (e.g., sensitive ecosystems/receptors, high emissions), and suggest a

more refined Tier 2 ERA process following USEPA ERA guidelines for reducing uncertainties. In addition, the public will be suspicious of the assumption that petroleum refinery emissions are not an ecological risk issue, so a more thorough justification is needed with site specific documentation.

A few strong contentions regarding ecological risks need more discussion and justification. For example, page 3-20 (second paragraph under 3.5.1) contends that if there are no adverse effects on humans, the "potential for adverse environmental effects…was considered to be insignificant. This assumption needs some scientific justification. Also, the rationale (e.g., pages 3-22 and J12 (J-3.2.4)) that no population effects will occur in a small water body because there are few individuals is flawed, and could be the reverse. Fewer individuals (lower abundance) means the population is more susceptible to extinction, particularly if there are few to no areas of refugia for recruitment of new individuals.

On page 2-29 it is stated that contaminant concentrations were evaluated against ecological benchmarks for sediment, soil and water. These comparisons were not found and must be reviewed. Which benchmarks (there are many for sediments)? What concentrations were used for each media? What was their associated uncertainty?

Cancer risk characterization: In Table 2-6 on page 2-22, the blank space for toluene cancer unit risk stands out because toluene has the greatest emissions. A person scanning this table would worry that this unknown could be a great source of risk. The document should explicitly discuss toluene cancer studies (from IRIS) and risk classification in the earlier section on dose-response. It should also provide a summary in the text here about the evidence being inconclusive, but since good studies were performed carcinogenic effects would likely have been observed if the risk were high. Thus, it cannot be dismissed, but there is no current evidence for significant concern.

In the discussion of uncertainties in dose-response relationships for cancer assessment, the most important uncertainty is probably that the upper bound is used for assessments.

The discussion of this issue in the last paragraph of page 2-32 is good. Page 2-33 describes the cancer guidelines accurately. However, with perhaps one exception, defaults were used. For example, on page 2-33, the paragraph dealing with pharmacokinetic models is accurate, but not relevant if none of the URE values were actually developed or modified through such an analysis. If they were not, this information should be deleted and only included in case studies where it was used, specifying the chemical for which it was used.

Characterization of chronic non-cancer risks: This section has about two full pages of description of the RfC/D methodology (pages 2-36 on). It should be reduced by about half, only providing information pertinent to understanding which uncertainties applied to the particular source category and which were accounted for in the RfC/D derivation. A full description of the methods is more appropriate for a separate general methods document. Some of the discussion here is redundant. A simple description will communicate the process better. The focus should be on the RfC, with a brief paragraph explaining where the RfD differs. Right now, much of the text treats the RfD as the "standard", when it doesn't apply (e.g., dosimetric adjustment). There are also a few missing elements or errors in the description of the methodology, which are described below.

a. Insert a sentence that states that the RfC has basic data requirements (e.g., at least a 90 day study, etc) before proceeding, to explain that an RfC is not guesswork.
b. Insert a short discussion about how many of the uncertainty factors (UFs) have redundant elements and therefore are conservative when multiple UFs are used. That is why EPA has a maximum of a total factor of 3000.
c. On page 2-37, under paragraph "1)", note that the heterogeneity UF includes children, people with preexisting disease, and other populations that may have added susceptibility. This is implied in the word "heterogeneity," but it is important to be direct about such a critical risk element.

d. On page 2-37, paragraph "2)" needs significant revision. The RfC methodology for extrapolation from animals to humans includes a dosimetric adjustment, resulting in the routine use of an UF of 3 to account for pharmacodynamic extrapolation. The RfD methodology does not do this routinely. This is a major difference. The paragraph implies that an UF of 10 is routinely used. Also, the paragraph talks about mg/kg/day which is only relevant to the RfD.

Characterization of acute health risks: The discussion on page 2-38 should be expanded to cover the uncertainties involved in the values chosen (e.g., AEGLs, ERPG). This is especially important since these are levels that cause effects, rather than "safe" levels. The discussion in this section should better parallel the section on chronic risks. The difficulty is that acute exposures did not really contribute much risk in the case study, but it still bears more discussion.

APPENDIX

Editorial suggestions for risk characterization sections:

1. As described in the EPA document, during the risk characterization step, information from other risk assessment steps is integrated to come to an overall conclusion about the risks involved. As a result, for the petroleum refinery discussion, Section 2.3.2 and the information in Tables 2-7 and 2-8 should be included in the Risk Characterization section. Similarly for the Portland cement discussion, Section 3.3 and the information contained in Tables 3-3 and 3-4 should be included in the Risk Characterization section. The details of EPA's cancer guidelines for early-life exposure (page 2-17, last paragraph) should be moved to section 2.2.6 on dose-response assessment. The risk characterization should provide some of this information, but delete the details for the age groups and also the BaP equivalence.
2. The risk characterization should "stand alone". For example, in some cases abbreviations are used excessively for the intended audience. All but very common abbreviations (e.g. HAP) should be avoided. For example, on page 2-19, "TOSHI" should be spelled out. The abbreviation URE should be defined on page 2-17, in paragraph 3.
3. On page 2-22, in Table 2-6, in the first row, specify that the URE is the upper bound (perhaps through a footnote).
4. The footnotes often provide excessive detail for the intended audience. Perhaps they could be summarized in plain English, with references provided for those seeking the precise words. Footnote 29 might be omitted.
5. P2-18ff. Section 2.2.7.2 Mixtures. P2-19, line 1. The word aggregate should be changed to cumulative since the intent here is to look at mixtures of different chemicals.
6. P 2-26 Table 2-8. Consider adding a footnote that defines the term "refined" used in the title.
7. P2-32, Section 2.4.2. Para 1 The description of durations not used could be deleted (i.e., just keep the descriptions for 1 hr and chronic durations).

8. P2-35. Last paragraph, "Chronic noncancer…" after the word represent, delete "chronic" and insert "70-year lifetime continuous exposure". Since everyone knows that such exposure scenarios are highly unlikely, the reader will automatically sense an uncertainty in the conservative direction.
9. P2-36 paragraph 3, line 3. Delete "relevant" and insert "sensitive" after endpoint. It is the "critical" endpoint, but such language doesn't really communicate effectively.
10. P2-38 para 1. Line 4. Insert "respiratory" before irritation.