UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

July 31, 2020

EPA-SAB-20-008

The Honorable Andrew R. Wheeler
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Consultation on EPA's Consolidated Human Toxicity Assessment Guideline

Dear Administrator Wheeler:

EPA's Science Advisory Board held a public meeting on June 23 - 24, 2020, and conducted a consultation with EPA staff on the Agency's proposed approach for developing a Consolidated Human Toxicity Assessment Guideline. Members of the Science Advisory Board's Chemical Assessment Advisory Committee also participated in the consultation.

The Science Advisory Board Staff Office has developed the consultation as a mechanism to provide individual expert comments for the EPA's consideration early in the implementation of a project or action. A consultation is conducted under the normal requirements of the Federal Advisory Committee Act (FACA), as amended (5 U.S.C., App.), which include advance notice of the public meeting in the Federal Register.

No consensus report is provided to the EPA because no consensus advice is given. Individual written comments were requested from all members of the Science Advisory Board and the Science Advisory Board Chemical Assessment Advisory Committee. The EPA's charge questions for the consultation are provided in Enclosure A. The individual written comments that were received from EPA Science Advisory Board members are provided in Enclosure B, and the individual comments that were received from members of the Science Advisory Board's Chemical Assessment Advisory Committee are provided in Enclosure C.

We thank the EPA for the opportunity to provide advice early in the Agency's process of developing a Consolidated Human Toxicity Assessment Guideline. In its charge to the SAB, the Agency has indicated that it plans to use a modular approach to develop the consolidated Guideline and has suggested that there be regular consultations with the SAB as Guideline modules are developed. Toxicity values are the foundation of many of EPA's activities and should have a sound scientific basis. The SAB strongly supports the suggestion of ongoing engagement with EPA staff and stands ready to provide advice to the EPA throughout the Guideline development process.

Sincerely,

/s/

Dr. Michael Honeycutt, Chair
EPA Science Advisory Board

/s/

Dr. Hugh A. Barton, Chair
SAB Chemical Assessment Advisory Committee

Enclosures (3)

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board (SAB), a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The SAB is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names of commercial products constitute a recommendation for use. Reports of the SAB are posted on the EPA Web site at http://www.epa.gov/sab.

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR
Dr. Michael Honeycutt, Division Director, Toxicology Division, Texas Commission on Environmental Quality, Austin, TX

MEMBERS
Dr. Rodney Andrews, Director, Center for Applied Energy Research, UK Research, University of Kentucky, Lexington, KY

Dr. Hugh A. Barton, Independent Consultant, Independent Consultant, Mystic, CT

Dr. Barbara Beck, Principal, Gradient, Boston, MA

Dr. Deborah Hall Bennett, Professor, Environmental and Occupational Health Division, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA

Dr. Frederick Bernthal, President Emeritus and Senior Advisor to the Board of Trustees, Universities Research Association, Washington, DC

Dr. Bob Blanz, Associate Director, Office of Water Quality, Division of Environmental Quality, Arkansas Department of Energy and Environment, North Little Rock, AR

Dr. Todd Brewer, Senior Manager, Grants, Education, and Utility Programs, American Water Works Association, Denver, CO

Dr. Joel G. Burken, Curator's Professor and Chair, Civil, Architectural, and Environmental Engineering, College of Engineering and Computing, Missouri University of Science and Technology, Rolla, MO, United States

Dr. Janice E. Chambers, William L. Giles Distinguished Professor and Director, Center for Environmental Health and Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Dr. John R. Christy, Distinguished Professor of Atmospheric Science and Director of Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL

Dr. Samuel Cohen, Professor, Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE

Dr. Louis Anthony (Tony) Cox, Jr., President, Cox Associates, Denver, CO

Dr. Alison C. Cullen, Interim Dean and Professor, Daniel J. Evans School of Public Policy and Governance, University of Washington, Seattle, WA

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

Dr. Susan P. Felter, Research Fellow, Global Product Stewardship, Procter & Gamble, Mason, OH

Dr. Joseph A. Gardella, SUNY Distinguished Professor of Chemistry, Department of Chemistry, College of Arts and Sciences, University at Buffalo, Buffalo, NY

Dr. John D. Graham, Dean, School of Public and Environmental Affairs, Indiana University, Bloomington, IN

Dr. John Guckenheimer, Professor Emeritus and Interim Director, Center for Applied Mathematics, Cornell University, Ithaca, NY

Dr. Margaret M. MacDonell, Department Head, Argonne National Laboratory, Lemont, IL

Dr. Robert E. Mace, The Meadows Center for Water and the Environment, Texas State University, San Marcos, TX

Dr. Clyde F. Martin, Horn Professor of Mathematics, Emeritus, Department of Mathematics and Statistics, Texas Tech University, Crofton, MD

Dr. Sue Marty, Senior Toxicology Leader, Toxicology & Environmental Research, The Dow Chemical Company, Midland, MI

Mr. Robert W. Merritt, Independent Consultant, Houston, TX

Dr. Larry Monroe, Independent Consultant, Braselton, GA

Dr. Thomas F. Parkerton, Senior Environmental Scientist, Toxicology & Environmental Science Division, ExxonMobil Biomedical Science, Spring, TX

Dr. Robert Phalen, Professor, Air Pollution Health Effects Laboratory, School of Medicine, University of California-Irvine, Irvine, CA, United States

Dr. Kenneth M. Portier, Independent Consultant, Athens, GA

Dr. Robert Puls, Owner/Principal, Robert Puls Environmental Consulting, Bluffton, SC

Dr. Kenneth Ramos, Executive Director, Institute of Biosciences and Technology, Texas A&M University, Houston, TX

Dr. Tara L. Sabo-Attwood, Professor and Chair, Department of Environmental and Global Health, College of Public Health and Health Professionals, University of Florida, Gainesville, FL

Dr. Mara Seeley, Unit Chief – Exposure Assessment, Environmental Toxicology Program, Bureau of Environmental Health, Massachusetts Department of Public Health, Boston, MA

Dr. Anne Smith, Managing Director, NERA Economic Consulting, Washington, DC

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

Dr. Jay Turner, Professor and Vice Dean for Education, Department of Energy, Environmental and Chemical Engineering, McKelvey School of Engineering, Washington University, St. Louis, MO

Dr. Brant Ulsh, Principal Health Physicist, M.H. Chew & Associates, Cincinnati, OH

Dr. Donald van der Vaart, Senior Fellow, John Locke Foundation, Raleigh, NC

Mrs. Carrie Vollmer-Sanders, Director, Agriculture Engagement Strategy, Efroymson Conservation Center, The Nature Conservancy, Indianapolis, IN

Dr. Kimberly White, Senior Director, Chemical Products and Technology Division, American Chemistry Council, Washington, DC

Dr. Mark Wiesner, Professor, Department of Civil and Environmental Engineering, Director, Center for the Environmental Implications of NanoTechnology (CEINT), Pratt School of Engineering, Nicholas School of the Environment, Duke University, Durham, NC

Dr. Peter J. Wilcoxen, Laura J. and L. Douglas Meredith Professor for Teaching Excellence, Director, Center for Environmental Policy and Administration, The Maxwell School, Syracuse University, Syracuse, NY

Dr. Richard A. Williams, Retired Economist (Food and Drug Administration), Independent Consultant, McLean, VA

Dr. S. Stanley Young, Chief Executive Officer, CGStat, Raleigh, NC

Dr. Matthew Zwiernik, Professor, Department of Animal Science, Institute for Integrative Toxicology, Michigan State University, East Lansing, MI, United States of America

SCIENCE ADVISORY BOARD STAFF
Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

U.S. Environmental Protection Agency
Science Advisory Board
Chemical Assessment Advisory Committee

CHAIR
Dr. Hugh A. Barton, Independent Consultant, Mystic, CT

MEMBERS
Dr. Richard Belzer, Independent Consultant, Mt. Vernon, VA

Dr. Tiffany Bredfeldt, Senior Toxicologist, Toxicology, Texas Commission on Environmental Quality, Austin, TX

Dr. Karen Chou, Associate Professor, Department of Animal Science and Environmental Science & Policy Program, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI

Dr. Harvey Clewell, Principal Consultant, Ramboll Environment and Health, Research Triangle Park, NC

Dr. Joanne English, Independent Consultant, Menlo Park, CA

Dr. David G. Hoel, Distinguished University Professor, Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, SC

Dr. Jacqueline Hughes-Oliver, Professor, Statistics Department, North Carolina State University, Raleigh, NC

Dr. Michael Jayjock, Exposure/Risk Assessor, Jayjock Associates LLC, Langhorne, PA

Dr. Wayne Landis, Professor and Director, Institute of Environmental Toxicology, Huxley College of the Environment, Western Washington University, Bellingham, WA

Mr. Charles McConnell, Director, Center for Carbon Management, Office of the Chancellor, University of Houston, Houston, TX

Dr. Dennis Paustenbach, President, Paustenbach and Associates, Jackson, WY

Dr. Isaac Pessah, Professor, Molecular Biosciences, School of Veterinary Medicine , University of California, Davis, Davis, CA

Dr. Joann Brooks Powell, Assistant Professor, Center for Cancer Research and Therapeutic Development (CCRTD), Clark Atlanta University, Atlanta, GA

Dr. Kenneth Ramos, Executive Director, Institute of Biosciences and Technology, Texas A&M University, Houston, TX

Dr. Thomas Rosol, Chair and Professor, Biomedical Sciences, Ohio Musculoskeletal & Neurological Institute, Ohio University, Athens, OH

Dr. Ted Simon, Toxicologist, Ted Simon, LLC, Winston, GA

Dr. Eric Smith, Professor, Statistics, College of Science, Virginia Tech, Blacksburg, VA

Dr. Chandrashekhar Sonwane, Fellow, Specialist Scientist, Aerojet Rocketdyne, Chatsworth, CA

Dr. Laura Vandenberg, Associate Professor, Environmental Health Sciences, School of Public Health and Health Sciences, University of Massachusetts - Amherst, Amherst, MA

SCIENCE ADVISORY BOARD STAFF

Dr. Suhair Shallal, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

Enclosure A

The EPA'S Charge Questions

SAB Consultation on EPA's Consolidated Human Toxicity Assessment Guideline

Charge to the SAB on the *Consolidated Human Toxicity Assessment Guideline*

Background

EPA has developed numerous guidelines and technical reports related to human toxicity assessment[1]. Some endpoint-specific toxicity documents were developed more than 2 to 3 decades ago (e.g., mutagenicity - 1986; developmental toxicity - 1991; reproductive toxicity - 1996; neurotoxicity – 1998). Since the development of these early toxicity guidelines, EPA has also developed additional guidelines that address common elements in Agency risk assessments, such as planning and scoping/problem formulation, and benchmark dose modeling. Many scientific advances have occurred since the development of the existing EPA guidelines; and there are also risk assessment elements and toxicity endpoints, such as immunotoxicity, for which EPA does not have guidelines. As a result, the Administrator tasked EPA's Risk Assessment Forum with revising existing or developing new assessment guidelines.

One of the early steps in this process was requesting advice from the EPA Science Advisory Board (SAB). This request was discussed with the SAB at a public meeting in June 2019, from which EPA received many valuable comments from SAB members. Having considered the comments from this SAB consultation[2], as well as internal Agency discussions, EPA is now initiating the development of a single Consolidated Human Toxicity Assessment Guideline ("Consolidated Guideline") that will focus on hazard characterization and dose-response assessment. Hazard characterization and dose-response assessment are two critical considerations which, when combined with exposure evaluation[3] in case- or location-specific circumstances, support risk assessment.[4]

EPA is proposing to revisit its overall approach to risk assessment guideline development. The Agency intends to utilize a modular approach in developing the Consolidated Guideline. This modular approach will result in the development of one consolidated guideline that consists of focused modules. This modular approach is similar to that taken by EPA in updating its Exposure Factors Handbook.[5] This contrasts with the past approach of developing discreet and independent toxicity-endpoint and common-element guidelines. Use of a modular approach in the Consolidated Guideline will allow EPA to accrue the benefits of consolidation, such as

---

[1] https://www.epa.gov/risk/risk-assessment-guidelines#tab-1

[2] https://yosemite.epa.gov/sab/sabproduct.nsf/357DC7E5C59BA9AD85258438005BA457/$File/EPA-SAB-19-003+.pdf

[3] See *Guidelines for Human Exposure Assessment* https://www.epa.gov/risk/guidelines-human-exposure-assessment

[4] See *EPA's Framework for Human Health Risk Assessment to Inform Decision making* (2014) https://www.epa.gov/risk/framework-human-health-risk-assessment-inform-decisionmaking

[5] https://www.epa.gov/expobox/about-exposure-factors-handbook

enabling EPA risk assessors to more easily access and use relevant parts of the Consolidated Guideline, while providing for an efficient and timely update of the Consolidated Guideline as modules are completed.

Given the number of commonalities in cancer and non-cancer assessments, the Consolidated Guideline will include assessment of both cancer and non-cancer endpoints. It will also include approaches that are common across endpoints and consideration of state-of-the-science approaches for characterization of dose-response, in addition to the incorporation of new approach methodologies (NAMs). Emphasis will be placed on examining the state-of-the-science and incorporating updated best practices for estimating risk at environmental exposure levels of concern for Agency decision-making.

The Consolidated Guideline will include two types of modules:

- Modules addressing common elements of an assessment (*i.e.*, "common-element" modules) that pertain to all health endpoints (*e.g.*, project planning and scoping, generic aspects of dose-response modeling), and
- Modules addressing specific types of toxicity ("endpoint-specific" modules) that focus on aspects of the hazard characterization and dose-response issues and methods that are specific to that toxicity-endpoint.

EPA will develop the Consolidated Guideline in a stepwise modular fashion (see page 6, Figure 1 illustrating the implementation approach). Modules will be developed and completed or updated individually in response to advances in science and Agency practice, without having to update entire sets of Agency guidelines. Any significant new aspects of the Consolidated Guideline will undergo public comment and external scientific peer review. EPA intends to complete the design of the Consolidated Guideline and prioritize the modules to be developed in December 2020. EPA will initiate the development of the modules in January 2021.

SAB Consultation

EPA considered the many recommendations submitted through the June 2019 SAB consultation, which particularly emphasized the need to update or add to EPA's risk assessment guidelines to ensure the use of the best available science at all phases of risk assessment and to provide the guidelines in a centralized location. Many SAB member recommendations were specific to toxicity endpoints and dose-response issues, including the need for updated guidelines on developmental toxicity, new guidelines on immunotoxicity, and considerations of dose-response issues, such as guidance for the use of various dose-response modeling approaches (e.g., model averaging), further consideration of the use of low-dose extrapolation approaches, additional consideration of endogenous production of environmental contaminants, and methods that would harmonize the evaluation of dose-response for cancer and noncancer effects. EPA considered these comments as the Agency developed the consolidated guideline concept.

This new consultation on the approach EPA proposes to use to develop the Consolidated Guideline is the first of what the Agency suggests should be regular consultations with the SAB during the development of this work plan and the many modules to follow. Consultation at this early stage is important because establishing a robust framework is key to developing a

Consolidated Guideline that will support EPA's use of the best available science in its risk assessments.

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- Use of various dose-response modeling approaches (e.g., model averaging);
- Further consideration of the use of low-dose extrapolation approaches;
- Additional consideration of endogenous production of environmental contaminants; and
- Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

Table 1: Proposed Modules

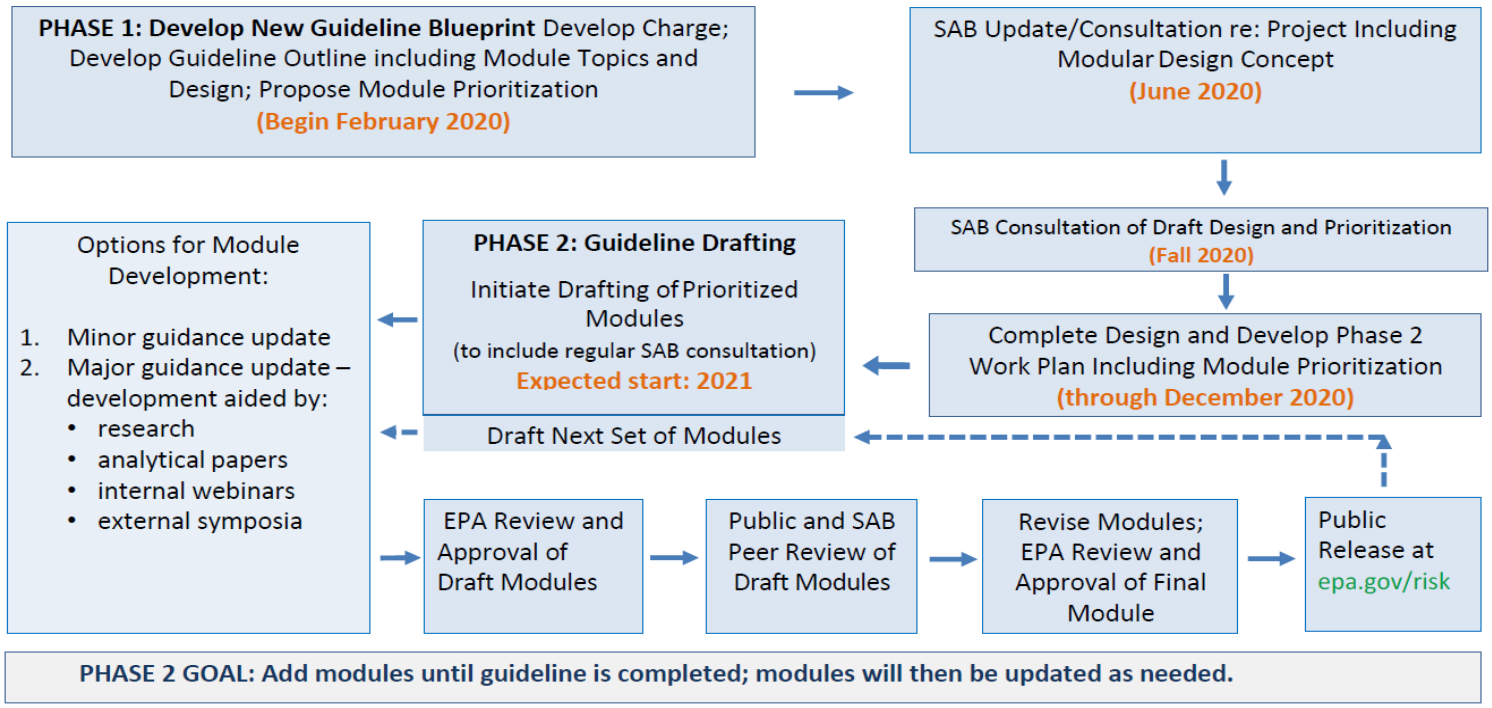
Modules are in order of how the Consolidated Guideline could potentially be organized, but not necessarily the order in which they would be written.

Common Element Modules *These proposed modules would address common elements of an assessment that pertain to all health endpoints*	**Module 1. Planning and Scoping a Human Toxicity Assessment** This module will provide an overview of human health toxicity assessment including key concepts such as fit for purpose, problem formulation, consideration of potential routes of exposure and overarching considerations including lifestage susceptibility, vulnerable populations and cumulative risk.

	Module 2. Identifying and Evaluating Toxicity Studies This module will cover general principles associated with collecting potentially relevant studies including conducting a literature search (systematic review), critically appraising different types of data (animal, epidemiological, chamber, modeling, in silico, NAMs, etc.) with respect to study design, power and reliability, data quality evaluation, and identifying data gaps.
	Module 3. Hazard Identification This module will cover integrating/weighing evidence/synthesizing results across studies, evaluating possible mechanisms/modes of action/adverse outcome pathways including human relevance, and consideration of lifestage susceptibility.
	Module 4. Dose-Response Assessment This module will cover a comprehensive set of issues including but not necessarily limited to: • Consideration of a unified approach for dose-response assessment; • Absorption, distribution, metabolism, and excretion (ADME) considerations; • Toxicodynamic versus toxicokinetic considerations; • Data quality considerations; • Types of dose-response data: animal tests; human chamber tests; epidemiological studies; occupational studies; high throughput testing; virtual tissue modeling; • Benchmark dose modeling including choosing a response rate, identifying a point-of-departure (POD) and extrapolation of dose-response to exposures lower than POD; • Deriving a POD, reference value, or margin of exposure; • Probabilistic modeling; • Model averaging; • Characterization of lifestage and population variability and vulnerability; • Physiologically Based Pharmacokinetic (PBPK) and Biologically Based Dose-Response (BBDR) modeling; • Use of adjustment factors including data derived extrapolation factors (DDEFs) and age-dependent adjust factors (ADAFs) to account for uncertainty, variability, susceptibility and use of generic default adjustment factors (e.g., body weight to the ¾-power); and • Cumulative risk considerations.

Endpoint Specific Modules *These proposed modules would focus on aspects of the hazard characterization and dose-response issues and methods that are specific to that endpoint*	**Module 5. Developmental Toxicity** **Module 6. Reproductive Toxicity** **Module 7. Immunotoxicity** *(no EPA guideline currently exists)* **Module 8. Carcinogenicity** **Module 9. Mutagenicity** *(mutagenicity as a mode-of-action would be addressed in both Module 3 – Hazard Identification & Module 4 – Dose-Response Assessment)* **Module 10. Neurotoxicity** **Module 11. Other Endpoints?** *(could add additional modules in the future for other issues or endpoints to potentially include, (e.g., Target Tissue Specific Considerations, Susceptible Lifestages and Population Groups)*	These proposed modules would cover definitions, critical concepts, test systems, data interpretation, and endpoint specific dose-response and exposure assessment considerations as needed.
Appendix	**Glossary** *(update after each module is developed)*	

Figure 1: Process/Timeline for Developing EPA's Consolidated Human Toxicity Assessment Guideline



Enclosure B

Individual Comments from Members of the EPA Science Advisory Board on EPA's Proposed Approach for Developing a Consolidated Human Toxicity Assessment Guideline

Dr. Hugh Barton B-2

Dr. Deborah Hall Bennett B-5

Dr. Janice Chambers B-7

Dr. Samuel Cohen B-9

Dr. Tony Cox B-12

Dr. Susan Felter B-13

Dr. Joseph Gardella B-19

Dr. Sue Marty B-20

Dr. Thomas Parkerton B-21

Dr. Robert Phalen B-23

Dr. Tara Sabo-Attwood B-24

Dr. Mara Seeley B-26

Dr. Kimberly White B-28

Dr. Richard Williams B-31

Dr. Hugh Barton

Discussion/Charge Questions for the Consolidated Human Toxicity Assessment Guideline

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

- The modular approach is appropriate as there are so many different aspects. From the outset, it needs to be defined what these guidelines are attempting to address. Historically, EPA human toxicity assessment guidelines focused on chronic or lifetime exposures rather than acute exposures, for example. With this modular approach, one could establish a framework that would be broader (e.g., including acute exposures such as accidental releases) that would be filled in over time, but in the meantime reference any current Agency guidance. Similarly, there have been differences in how toxicity assessments were done throughout the Agency under different laws, in different Offices of the Agency, and due to differences in available data. It is important to make clear what these guidelines are intended to address.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "<u>common element modules</u>" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

- The common element modules make sense as described as they represent the basic elements of toxicity assessment. In writing or updating these modules, they need to be open to new developments (e.g., new approach methods (NAMs)) and not lock in requirements for the whole animal studies that have been historically used. NAMs and in silico are mentioned in the described of Module 2 toxicity studies but need to be considered in each of these modules even though the methods for using them are still in development.

- Module 2 description: "chamber" is unclear, though in Module 4 it is more fully described as "human chamber tests".

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "<u>endpoint-specific modules</u>" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

- The addition of immunotoxicology guidance would be valuable and should be a high priority as this can underlie a host of human diseases.

- There is no guidance listed for most target organ toxicities (e.g., liver, kidney, spleen). At a minimum, this needs to be one module to address these or direct people to any existing Agency guidance.

- The proposed approach from EPA's Office of Pesticide Programs (OPP) for considering waivers for chronic/carcinogenicity studies includes assessment of genotoxicity, endocrine effects, and immunological effects as predictors for potential chronic or carcinogenic effects. Guidance for addressing endocrine effects is needed here.

- A challenge for these endpoint-specific modules is that NAMs and other approaches, such as toxicogenomic signatures evaluated in short-term animal studies, seem likely to be useful to evaluate the toxicity of a chemical but not necessarily be able to predict the endpoints or target organs that would be observed either in animals or humans. It may be too early to develop guidance for such approaches as this is an area of active research, but it could be identified as a module to be created in the future.

- Another challenge is that many human health effects important to public health are not predicted by in vivo animal toxicity studies. A road map for research and development efforts to address this is needed and some guestimate of a timeline for considering such effects in toxicity assessments developed. This might be a very short module but could be very informative.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

- The four common element modules are a reasonable first priority.
- A public commenter, Dr Fenner-Crisp, indicated that a mutagenicity MOA guidance was nearly complete, in which case that makes sense as a high priority to complete.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- Use of various dose-response modeling approaches (e.g., model averaging);
- Further consideration of the use of low-dose extrapolation approaches;
- Additional consideration of endogenous production of environmental contaminants; and
- Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

Harmonization of evaluation of dose-response for cancer and noncancer effects should be the highest priority. Further consideration of low-dose extrapolation approaches seems likely to be part of this. This task alone has multiple components.

- Outside chemicals acting through a few modes of action such as DNA-adducting mutagens or potent estrogens, it appears that tumors in animals are typically another chronic toxicity caused by toxicity processes that lead to a variety of chronic effects (e.g., histologically observable tissue damage). The historic differences in dose-response approaches has led to a focus on cancer endpoints to the detriment of endpoints, such as cardiovascular disease, that are also very important to human health.

Quantification of risk for cancer while continuing to estimate acceptable concentrations for noncancer endpoints has contributed to the under valuing of noncancer endpoints in risk assessments. Development of methods to estimate risks regardless of endpoint needs to be a high priority.

Dr. Deborah Hall Bennett

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

EPA has already developed an approach to identify and evaluate toxicity studies, specifically the approach they developed for the IRIS program, which has been reviewed by the National Academy of Sciences. They should continue to use this approach, rather than develop something new. They should refer to guidance already provided in multiple NAS reports on systematic reviews and hazard identification.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

There needs to be an effort to further understand human variation in both sensitivity to toxic effects and variability in exposure. There is increasing evidence of gene-environment interactions indicating differential sensitivity to exposures to particular compounds. Also, there is increasing evidence that those individuals suffering from a range of health disparities are more sensitive to exposures to compounds, with particles being a prime example. Those economically disadvantaged population that have health disparities also tend to have higher exposures to toxic compounds. A module to specifically consider both genetic susceptibilities and susceptibilities stemming from economic and health disparities needs to be included.

On variability in exposure, when considering exposures to things like a toxic waste site or factory, the old approach of considering exposures to, for example, the 95th percentile individual was perfectly adequate, as the exposed population was relatively small. However, as the EPA begins to consider compounds with widespread human exposure through either use in consumer products, or widespread water contamination from compounds such as PFAS, the size of the population exposed at above the 95th percentile becomes quite large, and thus variability in human exposure, and the ability to determine exposure to the highly exposed, needs to be considered.

Finally, a module on cumulative risk assessment needs to be developed, as the population is typically not exposed to a single compound, but rather a suite of compounds, many with similar modes of action.

All of these points were raised in the NAS report on Science and Decisions, and the recommendations of the NAS should be followed.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See

Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

There should be a module looking at endocrine related endpoints. The EPA has dedicated a significant amount of resources into developing methods to determine if compounds are endocrine disruption compounds, and if they are, through which endocrine system they operate. This should be brought in to a module looking at the impact of low-dose exposures to this class of compounds.

There should also be a module looking at cardiovascular endpoints. The scientific community has extensively studied the impact of air pollution and other compounds on cardiovascular health, and thus there is a rich literature to draw on to develop an endpoint specific module in this area.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

The NAS report on Science and Decision has made clear that there are often toxic effects at low dose, and has proposed recommendations for a unified dose response approach. The EPA should follow the recommendations of the NAS.

Dr. Janice Chambers

Charge Questions for Human Toxicity Assessment

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

The modular approach makes sense in that it will be easier to concentrate on revision of each section in a focused manner and it will be easier to revise individual modules when needed and replace modules than the entire guidance document. It will also be more efficient to gain SAB advice on the updates by having a more focused approach and group of scientists to advise with individual topics. The timeframe presented in Figure 1 is probably optimistic, especially if substantial rewrites or revisions are needed for some of the modules.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Probably the list is complete although as you work through these issues, especially modules 2 and 4, additional topics may become apparent. Some of the approaches listed, e.g., in silico and NAMs, are new and will require more thought and vetting than the more traditional data sets. Some of the types of data that need guidance on their utility, e.g., epidemiology, will require more input and consideration than others. I use the example of epidemiology because there are serious concerns about some of the epidemiology studies that have been published—we had an SAP about this several years ago when I was on the SAP and I am not sure that the meeting really resolved issues very well. Some epidemiology studies may be considered "valid" in that the math used to come up with the associations was done correctly, but there is a tendency for those conducting the studies to use these associations as evidence of causation, and that is not likely to be true, especially if the study has not considered comorbidities and other confounders effectively. Should that be the case, the study may not be considered suitable for regulatory purposes, and good guidance is essential in this area. Epidemiology studies definitely cannot provide evidence of causation with any of the accuracy that controlled animal studies provide, and this topic needs more critical appraisal than I saw from the earlier SAP meeting. Another topic that will require considerable good guidance will be cumulative risk considerations, since there are so many factors that can come into play (if the intent is to go beyond common mechanism of action of chemicals for factors that contribute to cumulative risk) and if such hard-to-quantify factors such as socioeconomic status (SES) are to be considered in cumulative risk considerations; precise definitions on what is considered and what is not considered in cumulative risk need to be developed. I would strongly suggest that EPA staff solicit scientific input from SAB members and

other scientists who have expertise in some of the newer and/or more complex topics in order to develop the guidance.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

One category missing from the Endpoint Specific Modules group is organ system specific toxicities, such as liver, kidney, and lung.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Module 1 is probably quite straightforward and could be updated rather quickly, so would be a good place to start. The others will require more thought and discussion.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

• Use of various dose-response modeling approaches (e.g., model averaging);
• Further consideration of the use of low-dose extrapolation approaches;
• Additional consideration of endogenous production of environmental contaminants; and
• Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

The third option above is the lowest priority because it is specific to only a relatively few toxicants.

Dr. Samuel Cohen

Summary of Recommendations from June 20, 2019, Consultation with Members of the EPA Chartered Scientific Charter SAB and CAAC

I strongly endorse the efforts by the EPA to update their guidance for overall risk assessment approaches, especially their attempts to unify the cancer and non-cancer risk assessment. This is particularly true for nongenotoxic chemicals. Some specific comments regarding the points listed in the document sent to us follow.

Under problem formulation and scoping, I believe that the last bullet point, "reality check," is particularly important. This has become quite evident in recent assessments, such as ethylene oxide, and others.

Under harmonization, I strongly support the effort to harmonize guidelines for cancer and non-cancer effects, including the dose response. This should be especially true for nongenotoxic chemicals (see below regarding genotoxicity assessment). Since for nongenotoxic chemicals, the mode of action always includes a precursor key event that is a non-cancer toxicity, protecting against this non-cancer toxicity will also protect against the risk of cancer. In particular, the default assumption for nongenotoxic carcinogens should be a threshold, nonlinear extrapolation to low dose, similar to what is performed for other types of toxic endpoints. Since the precursor lesions will be other types of toxicity beside cancer, the approach for non-cancer and cancer can be entirely the same. This requires that there be some understanding of mode of action, but again, it is essential that for nongenotoxic chemicals the default assumption be that there is a threshold. The continued use of a linear, non-threshold extrapolation to low dose is biologically inappropriate. Also, I would strongly encourage the EPA to utilize descriptors rather than just a scoring or labeling approach. The descriptors are much more useful in a risk management setting. For example, if the toxicity occurs only at a dose above a threshold that leads to a specific toxicity, there is no toxic risk, including cancer risk, below that level. Thus, if there is no evidence of the toxic endpoint precursor, there is no risk of cancer.

Under the general cancer issues, there are several issues that need to be addressed. Although there need to be updates of the cancer guidelines regarding statistical methods, it is important to emphasize that the biology is the predominant determinant of the risk assessment, not the statistical approach. For example, the standard joke regarding causation versus association illustrates this point strongly. One night a drunk goes out and drinks several scotch and sodas and gets a terrible hangover, becomes very sick. So, the next night, he goes out and has bourbon and soda, and the same thing happens. The third night, he goes out and has rye whiskey and soda, and the same thing happens. When he wakes up the third morning, he is terribly sick and he says, I have to just stop drinking that soda, it's making me sick. It is a 100% correlation, but biologically ludicrous. Although, we laugh at this, there are numerous examples in the literature from epidemiology studies that make this mistake. There appears to be an increasing emphasis for Bayesian analysis. This might be helpful in some instances, but does not serve as a panacea for solving statistical issues. You still have to have basic biological information to make the judgements, both with regard to relevance and with regard to dose. Again, I would emphasize that the linear-no-threshold (LNT) approach as a default for low-dose

extrapolation is totally inappropriate, certainly for nongenotoxic chemicals. As indicated above, the default assumption for nongenotoxic chemicals should be a nonlinear, threshold approach. With regard to animal models, it is important to keep in mind the relevance of the model being used, and especially the relevance of the mode of action for human risk. Likewise, the relevance of the dose at which the toxic endpoints are identified needs to be addressed. Careful consideration for MTD and KMD is especially important for extrapolating to lower doses. If toxicity is only seen at doses above the MTD or above the KMD, these are not appropriate for consideration for risk assessment. This should be explicitly stated in the guidelines. The suggestion to convene panels for human relevance of certain animal tumors is critical at this time. There remain several animal rodent tumors and modes of action that continue to be considered relevant to humans which are not actually relevant either qualitatively or quantitatively. These panels should include experts from veterinary and human medicine in addition to toxicology, pathology, statistics, and molecular biology. With regard to NAMs, I encourage the agency to continue development in this area, but I also caution that reasonable biological principles continue to be incorporated into these attempts. For example, doses used in these studies should not be above the MTD or above the KMD. Findings above those doses are meaningless with regard to actual human risk. In addition, the relevance of specific toxic endpoints in animal models needs to be addressed. This has become increasingly obvious in the pharmaceutical industry, where approximately one half of the pharmaceuticals that have been tested in two-year bioassays have positive results, and yet are still used in medicine. Examples include the statins (rodent liver tumors), proton pump inhibitors (gastric neuroendocrine tumors), and fibrates (PPARα activators). These models are completely irrelevant to humans, based not only on biological evaluations, but extensive epidemiology studies involving hundreds of thousands of individuals. There are actually very few rodent tumor models that are relevant to humans. Likewise, there are several toxic endpoints that occur in animal models that do not extrapolate to the human situation.

With regard to specific cancer issues, there are several that I just listed. In addition, some of the specific points that are listed here need to be addressed. One that is critical is the bar for mutagenic MOA. There needs to be some clear guidance provided with regard to interpretation and consideration of the numerous genotoxicity assays that are performed. Utilization of OECD guidelines in this analysis, as well as the quality of specific studies needs to be carefully addressed. There are way too many examples of positive results in the literature that are not reproducible or that only occur under circumstances that do not extrapolate to the whole organism. A specific statement should be made that a negative finding in an *in vivo* assay overrides the findings of a positive result in an *in vitro* assay. With regard to cell-proliferation requirements, there should be some mention that a labeling index (such as BRDU, Ki-67, or PCNA) needs to be included for *in vivo* studies, particularly in short term studies, since reliance on histopathology will not be adequately sensitive. The suggestion to reevaluate practices for determining statistical significance for common tumors is essential. This was described originally by Joe Haseman at the NTP, and has been adopted by OECD and by FDA. There is strong biologic as well as statistical support for this approach. Without defining this, and even requiring it, leads to way too many false positive results from the bioassay. The suggestion to develop guidance for use of initiation-promotion studies for cancer I believe is misguided. The initiation-promotion model is outdated, and generally can be translated to initiation being synonymous for genotoxicity and promotion being for increased cell proliferation. The reality is

that chemicals that act as initiators or promotors are actually carcinogens when investigated in the full two-year bioassay. The only advantage of using this model is that it can identify a nongenotoxic carcinogen in a shorter time, but the same information can be garnered by even shorter term cell-proliferation studies. In addition, this model does not help in addressing the issue of relevance to human cancer risk of the tumors that are induced. I would strongly encourage the EPA to abandon any consideration of the initiation/promotion studies.

Samuel M. Cohen, MD, PhD
Professor, Department of Pathology and Microbiology
University of Nebraska Medical Center

Dr. Tony Cox

Comments in response to the charge questions for the SAB consultation on EPA's Human Toxicity Assessment Guideline.

- Validation of dose-response models and characterization of model uncertainty should be addressed in detail in Module 4 (Dose-Response Assessment).
- Chronic inflammation and inflammation-related MOAs should be added, either as a separate module, or as a distinct part of Module 7 (Immunotoxicity). Elucidation of the role of inflammasomes (especially the NLRP3 inflammasome) in many exposure-related diseases has revolutionized biological understanding in recent years, and this should be reflected in biologically based and biologically motivated toxicity assessment and risk assessment.
- Bayesian networks, causal biological network models, and systems biology methods and models should be added to Module 4.
- Ensemble methods other than model averaging (e.g., individual conditional expectation plots) should be added to Module 4.

Dr. Susan Felter

EPA SAB Consultation: Human Toxicity Guidelines

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

In general, I am very supportive of the EPA's plan to use a modular approach to develop its Consolidated Guideline and to consider two types of modules: "common element" modules, which apply across all endpoints and "endpoint-specific" modules that update and expand existing guidelines (or develop new ones) that address specific types of effects.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

The "common element modules" include:
Module 1. Planning and Scoping a Human Toxicity Assessment
Module 2. Identifying and Evaluating Toxicity Studies
Module 3. Hazard Identification
Module 4. Dose-Response Assessment

Module 1 ("Planning and Scoping"): I am glad to see "Planning and Scoping" as the first critical consideration in any risk assessment and note that this is consistent with the NAS (2009) "Silver Book" framework for risk-based decision-making. The current description of Module 1 recognizes potential routes of exposure. EPA should consider expanding this to include *duration* and *magnitude* as well. Historically, EPA's risk assessment guidelines (RAGs) have focused on chronic exposure. Does the Agency anticipate expanding this to include guidance for shorter-term exposures? Magnitude of human exposure is also critically important to consider *upfront* as this will help prioritize the extent to which a screening assessment (e.g., based on the TTC, the Threshold of Toxicologic Concern) might be sufficient, or the extent to which refinement of a chemical-specific risk assessment is appropriate (for example, using a PBPK model in the assessment).

Module 2 ("Identifying and Evaluating Toxicity Studies"): EPA should consider expanding this (or adding another module) beyond toxicity studies to include information on physical-chemical properties that will be important to consider, and other data that will be important in the risk assessment but would not necessarily be classified as "toxicity studies." This would likely include some of the NAMs as well as more traditional PK studies, MOA studies, etc. If the intent is for Module 2 to be the step where all relevant data are identified/collected/evaluated, then a broader title would be appropriate (e.g., "Identifying and Evaluating Studies/Data Relevant for Risk Assessment").

Modules 3 (Hazard Identification) and 4 (Dose-Response Assessment): Historically, Hazard Identification (HI) and Dose-Response Assessment (DRA) have been treated as separate steps, starting with the NAS (1983) paradigm and continuing today (although I note that NAS (2009) put these 2 elements together in one box). Our understanding of toxicology today is such that, in fact, these should *not* be separated and that HI is only appropriate in the context of DRA. EPA's description of Module 3 (HI) indicates that it will include "evaluation of possible mechanisms/modes of action/adverse outcome pathways including human relevance…" I think this offers another argument to no longer consider HI separately from DRA since we know that the MOA can be different at a low dose than a high dose, and the 'hazard' can be different depending on the route, magnitude, and duration of exposure. An example is EPA's risk assessment in IRIS for 2-butoxyethanol, for which the summary states, "Under the Guidelines for Carcinogen Risk Assessment (U.S. EPA, 2005), EGBE is deemed "not likely to be carcinogenic to humans at environmental concentrations at or below the RfD and RfC, based on laboratory animal evidence, mode-of-action information, and limited human study information." In this example, it is clear that the dose-response assessment had to be considered to help inform that HI statement. Finally, it is not clear how (if) HI would be addressed with NAMS that might identify a biological threshold for activity/toxicity, but not necessarily be associated with a particular endpoint. One option might be to change Module 3 to focus on evaluation of data that feed into the eventual quantitative risk assessment (e.g., to mechanistic/MOA data) and then have Module 4 be "Quantitative Risk Assessment"

Specifically for "Dose-Response Assessment," I offer the following considerations:

- Consider changing to "Quantitative Risk Assessment" since we will likely be dealing with new types of data that go well beyond the traditional DRA associated with evaluation of apical endpoints in a rodent toxicity study.
- I fully support a unified approach based on the underlying biology – if, for example, tumors develop secondary to a sustained toxic insult, the quantitative risk assessment should focus on ensuring that the public is protected against that toxic insult, such that all secondary effects are also protected against. This suggests that the default approach for nongenotoxic carcinogens should also be thresholded, as it is for the RfD and RfC.
- There should always be the flexibility to include mechanistic/MOA data in considering the best approach to quantitative risk assessment. This is true for genotoxic/DNA-reactive substances as well as nongenotoxic as we continue to learn more about thresholds for genotoxicity (e.g., Hartwig et al., 2020; Metruccio and Moretto, 2018; Jenkins et al., 2010; Müller et al, 2009).
- It is increasingly recognized that toxicokinetics (ADME) is a critical part of any robust risk assessment. This includes consideration of species-specific TK and also the impact of *dose* such that nonlinearities in ADME should be taken into account. Effects seen at very high doses that have saturated relevant TK parameters may have no relevance to human health, but still lead to classification (hazard ID), and can even drive the quantitative risk assessment. These findings can also lead to extensive follow-up studies and evaluations that are costly in terms of animals, time, and money. This argues for inclusion of the KMD (Kinetically-derived maximum dose) both in designing any new experiments, and also in the evaluation of existing data.
- Beyond consideration of nonlinearities in TK that are important for assessing dose, it is also

important to consider what is known (or not) about human exposure (this ties into my suggestion that this be considered *up front* in the "Problem Formulation" stage). This is especially important as protocols are being developed for NAMs. I think we will want to avoid a situation where testing (*in vivo or in vitro*) is conducted at non-physiological doses to which human exposures would never come close. Testing at inappropriately high doses could also result in findings that then require extensive follow-up with no benefit to human health and, in fact, the possibility of negative (unintended) consequences. As an example, I recently published a case study on β-myrcene (Felter et al., 2020), where the doses administered in the rodent cancer bioassay (including the lowest dose) were five-six orders of magnitude higher than human exposures. A finding of an increase in tumors resulted in a challenge to the FDA to remove β-myrcene (a flavor substance found naturally in many foods) from the list of approved food additives because it is now considered to be 'a carcinogen' even though no regulatory agency (including the FDA) has concluded that there is any safety concern associated with human exposure to β-myrcene as a flavor substance.

- Regarding benchmark dose (BMD) modeling, I think model-averaging is appropriate, but consideration of underlying biology should be considered and should always take precedent when choosing which model(s) to include. Where this is not possible, one could default back to averaging of the current suite of models used by EPA. The same is true for Bayesian methods. They are often described in a way that suggests increased confidence in a risk assessment, but I am not convinced this is true, and suggest that the only way we can really increase our confidence is through a better understanding of the underlying biology/toxicology.
- EPA has started to use a BMD approach called "MS-Combo" in which all tumor types are being included in the quantitative evaluation (vs. modeling the most sensitive endpoint). This should not be done without very careful consideration of what is being modeled, including assurance that the tumors are *independent* of each other. I believe EPA had this tool reviewed by 3 experts many years ago, but it's not clear that all of the recommendations are being followed in the implementation of this software. As with other aspects of cancer risk assessment, it is critical that biological considerations are put ahead of statistical ones.
- Significant work is still needed to support quantitative use of NAMs in risk assessment. For example, if toxicogenomics data are available, how can/should they be integrated into the risk assessment? Would the same UFs be applied to a 'genomic no response' level (or other POD from a NAM study)? Consideration should be given to approaches described by Yauk et al. (2020), Cheung et al (2018).
- Does EPA want to develop guidance to help with the *design* of NAMs studies specifically to increase their utility for risk assessment? This would especially include consideration of dose/concentration to avoid generating data that are not relevant to human exposures, rather than basing the highest concentration on something like solubility or lethality (to cells).
- Guidance for *in vitro* to *in vivo* extrapolation will be needed to enable the use of *in vitro* data for quantitative risk assessment.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

EPA has identified 6 "endpoint-specific modules": Developmental Toxicity; Reproductive Toxicity; Immunotoxicity (new); Carcinogenicity; Mutagenicity; and Neurotoxicity. Each of these endpoints warrants guidance for risk assessment. It will be helpful if the guidance can provide an overview of the known MOAs for each endpoint and how these should be considered in the conduct of a risk assessment. For carcinogenicity, it will be important to include guidance for establishing a mutagenic vs non-mutagenic MOA, including recognition that this can be dose-dependent (e.g., see Hartwig et al., 2020). Many *in vitro* genotoxicity studies conducted under current protocols may yield a positive result, but this does not equate to a mutagenic mode of action. Increased consideration should be given to the potential for thresholds in carcinogenicity, including for genotoxic substances. For developmental toxicity, it is important to provide guidance regarding the consideration of maternal toxicity.

As toxicogenomics (and other NAM) data are increasingly available, it will be important for EPA to establish guidance for the use of these data in risk assessment, even when the target organ is not known. This also applies to data such as ToxCast that are being used for "Hazard Identification" (e.g., the Key Characteristics of Carcinogens) – while this might be a helpful tool to consider how NAMS data can help elucidate the MOA for a substance, it is also fraught with challenges and offers the potential for significant mis-use. Guidance for interpretation/use of ToxCast data in risk assessment is a high priority given that the data are publicly-available.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

It seems logical to start with the common modules as they will inform the process for the endpoint-specific modules. For the endpoint-specific modules, the priority can be considered based on what endpoints are most impactful for EPA risk assessments, including high-profile assessments that have been controversial. Carcinogenicity will likely be at/near the top of the priority list.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

Of the issues listed, I consider the highest priority to be further consideration of the use of low-dose extrapolation approaches and methods that harmonize the evaluation of dose-response for cancer and noncancer effects. For low-dose extrapolation, this is most important for cancer risk assessment, where current default approaches (including for nongenotoxic carcinogens where the MOA is not known) result in extrapolation over ~ 5 orders of magnitude and thus will generally drive any risk assessment. Much has been learned since the early days of quantitative cancer risk

assessment when EPA made a policy decision to use (initially) the linearized multistage model, and subsequently, linear extrapolation from a POD such as the $BMDL_{10}$. Even for substances considered to be "genotoxic carcinogens," updated guidance should recognize and provide flexibility for approaches not limited to a linear model based on some evidence for genotoxicity which might be at a high dose only and not relevant to exposures encountered by humans (see, for example, Hartwig et al., 2020). For nongenotoxic carcinogens, this point is even more important as there is currently a very high bar for assessments in IRIS to be based on anything other than linear low-dose extrapolation. To my knowledge, all major regulatory agencies outside of the U.S. have adopted a threshold-based approach as the default for nongenotoxic carcinogens, without a requirement to fully elucidate the MOA.

After low-dose extrapolation and harmonization of dose-response for cancer and noncancer effects, I think additional consideration of endogenous production of environmental contaminants should be the next priority (e.g., see Andersen et al., 2019; Hartwig et al., 2020). Consideration should also be given to information that can be gleaned from human exposure to naturally occurring substances in a healthy diet (e.g., Autrup et al., 2020).

One I would *add* to the list is increased consideration of human relevance of a number of rodent tumors, the mouse liver, mouse lung, and rat kidney being at the top of the list. While a number of workshops have been held over the years related to mouse liver tumors, no change related to EPA's cancer risk assessment guidelines has come from this and mouse (usually B6C3F1) liver tumors are often the only evidence of an increase in tumors that drives a cancer risk assessment, still defaulting to linear low-dose extrapolation. It is noteworthy that guidance from the EU classification system for carcinogenicity [1] (now replaced with the GHS) stated that, "if the only available tumour data are liver tumours in certain sensitive strains of mice, without any other supplementary evidence, the substance may not be classified in any of the categories" and "particular attention should be paid to cases where the only available tumour data are the occurrence of neoplasms at sites and in strains where they are well known to occur spontaneously with a high incidence." The difference in approaches is striking and suggests that this is an important topic for re-evaluation. It is a similar situation for the rat kidney (e.g., Hard et al., 2013) and mouse lung (e.g., Cohen et al., 2020) as well as many other tumors that occur with a high spontaneous frequency in different strains of rats and mice.

References

Andersen ME et al. 2019. Considerations for refining the risk assessment process for formaldehyde: Results from an interdisciplinary workshop. Regul Toxicol Pharmacol, 106: 210-223. https://doi.org/10.1016/j.yrtph.2019.04.015

Autrup, H. et al. 2020. Human exposure to synthetic endocrine disrupting chemicals (S-EDCs) is generally negligible as compared to natural compounds with higher or comparable endocrine activity. How to evaluate the risk of the S-EDCs? Toxicology Letters, 331: 259-264. https://doi.org/10.1016/j.toxlet.2020.04.008

[1] https://ec.europa.eu/environment/archives/dansub/pdfs/annex6_en.pdf

Cheung, C. et al. 2018. Evaluation of the Use of Toxicogenomics in Risk Assessment at Health Canada: An Exploratory Document on Current Health Canada Practices for the Use of Toxicogenomics in Risk Assessment. https://www.canada.ca/en/health-canada/services/publications/science-research-data/evaluation-use-toxicogenomics-risk-assessment.html

Cohen S., et al. 2020. Relevance of mouse lung tumors to human risk assessment. Journal of Toxicology and Environmental Health, Part B. https://doi.org/10.1080/10937404.2020.1763879

EU. ANNEX VI. General Classification and Labelling Requirements for Dangerous Substances and Preparations. https://ec.europa.eu/environment/archives/dansub/pdfs/annex6_en.pdf

Felter, SP. et al. 2020. How the 62-year Old Delaney Clause Continues to Thwart Science: Case Study of the Flavor Substance β-myrcene. Regul. Toxicol. Pharmacol., Jun 6;115:104708. doi: 10.1016/j.yrtph.2020.104708. https://pubmed.ncbi.nlm.nih.gov/32522581/

Hard, GC et al. 2013. Consideration of Rat Chronic Progressive Nephropathy in Regulatory Evaluations for Carcinogenicity. Toxicol Sci., 132(2): 268–275. https://doi.org/10.1093/toxsci/kfs305

Hartwig, A. et al. 2020. Mode of action-based risk assessment of genotoxic carcinogens. Archives of Toxicology (2020) 94:1787–1877. https://doi.org/10.1007/s00204-020-02733-2

Jenkins, GJS. et al. 2010. Genotoxic thresholds, DNA repair, and susceptibility in human populations. Toxicology 278: 305–310. https://doi.org/10.1016/j.tox.2009.11.016

Mettrucio, F. and Moretto, A. 2018. Genotoxicity in risk assessment: is it time to use a threshold approach? Current Opinion Toxicol, 11-12:21–26 https://doi.org/10.1016/j.cotox.2018.11.007

Müller, L. et al 2009. Ethyl methanesulfonate toxicity in Viracept-A comprehensive human risk assessment based on threshold data for genotoxicity. Toxicology Letters, 190(3):317-329 DOI: 10.1016/j.toxlet.2009.04.003

Yauk, CL et al. 2020. A cross-sector call to improve carcinogenicity risk assessment through use of genomic methodologies. Regul Toxicol Pharmacol, 110: 104526. https://doi.org/10.1016/j.yrtph.2019.104526

Dr. Joseph Gardella

With respect to the development of modules, I think the development of a core module for an emerging pollutant category like PFAS/PFCs would be timely and useful in supporting the work EPA is doing presently to develop analytical methodologies, establish values for water quality measurements and of course give more insight in to human toxicity assessment for this important and immediate need for information on these chemicals.

Dr. Sue Marty

I do not have many comments related to the Consolidated Human Toxicity Assessment Guideline as the information seems like a high-level overview.

Question 1: I think that a module approach for developing the Consolidated Guideline seems reasonable, but the EPA will need to review how each module fits into the overall guideline to ensure consistency in their approach.

Question 2: The modules listed in Table 1 seem appropriate (I did not identify omissions) and the descriptions were sufficient for the purposes of this document. Overall, the stepwise approach proposed by the EPA seems logical.

Question 3: For the endpoint specific modules, is endocrine included in the reproductive or developmental endpoints? If so, where do thyroid-related effects fit?

Question 4: With respect to sequence, the modules appear to be laid out in a logical order. Once the elements of modules 1-4 are in place, there will be some consistency in the review and application of data, that will facilitate the EPA's risk assessment procedures, regardless of the specific endpoint.

Question 5: For dose-response models, the current state of the science generally supports that linear low-dose assessments for carcinogens, particularly non-DNA-reactive compounds, are overly conservative. Generally, other specific endpoints are recognized to have thresholds as well. It would be useful to address this in EPA's Consolidated Approach. I look forward to updates on the development of these modules.

Dr. Thomas Parkerton

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

RESPONSE: EPA's proposed modular approach is a sensible way to tackle the issue of human health risk assessment. The modular approach will allow each module to be reviewed and updated as needed in a more rapid fashion. However, it is unclear how the proposed initiative effectively contributes to the Administrator's goal of reducing and eliminating animal testing. It is recommended that EPA consider how the consolidated guidelines can be developed and the modular elements of the guide structured and prioritized to address this key priority.

In EPA's Charge to the SAB on the Consolidated Guideline, it is stated that as this initiative proceeds, regular consultations with the SAB is envisioned to ensure a robust framework is developed to support EPA's use of the best available science in its risk assessments. However, given the extent of effort that is planned, I question if SAB advice to EPA for the Consolidated Guideline may be better served through the establishment of a dedicated SAB subpanel rather than the Charted SAB. A key advantage of this alternate approach for SAB engagement is that a broader array of subject matter experts covering the technical aspects of both common and endpoint specific modules could be assembled from different sectors to offer timely expert input to EPA on this ambitious endeavor. Relevant experts from the Chartered SAB could be included on this subpanel to facilitate dialogue with the broader Chartered SAB members as needed.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

RESPONSE: It is suggested that module 1 include non-testing approaches, like Threshold of Toxicological Concern (TTC), for use as first step to determine the need for more in depth hazard evaluation. Additionally, for this module, how can hypothesis-based prioritization be employed to logically focus the need for more detailed assessment? EPA should also consider describing best practices to help improve the replicability and transparency of hazard related research based on the recent National Academies report[1] and the Center for Open Science platform to publish experimental protocols a priori; share data, materials, or code; and increase collaboration between investigators[2]

[1] Committee on Science, Engineering, Medicine, and Public Policy; Policy and Global Affairs; National Academies of Sciences, Engineering, and Medicine National Academies of Sciences, Engineering, and Medicine 2019. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303

[2] Center for Open Science. https://www.cos.io/

For modules 2 and 3, it is recommended that EPA incorporate state of science approaches for data collection, quality scoring, systematic review, weight of the evidence analysis and mode of action assessment. A framework for deciding when New Approach Methods (NAMs) are deemed reliable to include in WoE evaluation would be valuable to incorporate or provide as a separate module. Module 2 should also cover reporting and analysis of uncertainties in toxicity test data as well as uncertainty and confounding factors in epidemiology studies. EPA may also want to consider additional common modules on the identification and analysis of subpopulations (as assumptions about population susceptibility are often applied) as well as hazard communication.

It is suggested that Module 4 be divided into threshold and non-threshold dose-response models. Presumably, threshold models would cover the majority of endpoints. Focus should be on the process and the methodological considerations that guide dose response assessment and how additional, targeted, fit-for-purpose dose-response data can reduce uncertainty in risk assessment.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

RESPONSE: It is important to clarify the purpose for these modules, i.e. how to interpret data generated from studies focusing on these endpoints or identify how risk assessment is done for these endpoints. If the later, it seems redundant with the dose response module. In the former, a key issue will be how are NAMs to be incorporated? Further, as NAMs are rapidly evolving, how will the guidance be practically updated?

For the endpoint specific modules it is suggested to consider the use of a decision tree or flow chart similar to the one found in Organisation for Economic Co-operation and Development (OECD) TG 150. The OECD TG 150 discusses the use of in vitro assays to detect potential endocrine disruption. In this guidance, there are decision trees describing the potential interpretations of the in-vitro tests and then what next test the researcher should consider and help make the connection between mechanistic, in-vitro and in-vivo data.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

See response to charge question (1). It is recommended that priorities be guided by the overarching goal of reducing and eliminating animal testing.

Dr. Robert Phalen

As an overview observation on EPA's Human Toxicity Assessment Guideline, and New Approach Methods and Reducing the Use of Laboratory Animals for Chronic and Carcinogenicity Testing, the other SAB members' comments thus far are insightful and relatively thorough. I have little to add to them, so I will restrict my comments to where I might contribute new ideas for the EPA to consider.

EPA's Human Toxicity Assessment Guideline:

Charge Question 1, Figure 1.

Module 1. A consideration of potential significant adverse health related trade-offs of regulatory actions could be added to the Planning and Scoping considerations. For example is the chemical under consideration uniquely important for protecting crops, livestock and/or human populations; and for which there are no viable substitutes. For example, control of crop damage leading to increased carcinogenicity or decreased nutritional content should be taken into account.

Module 3. The EPA might consider adding "Benefit Assessment" as a potential negative/offsetting relevant hazard.

Module 4. Current absorption, distribution, metabolism, and excretion (ADME) models may not adequately assess the long term storage (as in body fat) and future release (as in later weight loss) that could lead to significant re-exposure and adverse health effects.
Lipid soluble chemicals that are not harmful to fatty tissue but are neurotoxic are an example.

Module 11. Behavioral endpoints, including learning disorders should not be neglected.

Dr. Tara Sabo-Attwood

Consolidated Human Toxicity Assessment Guideline

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

The modular approach is appropriate and working through modules seems like an effective way to prioritize and revise the workflow. If the end game here is risk/safety assessment then exposure assessment seems to be missing as a stand-alone module.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Vulnerable populations will need to be clearly defined across the modules as there are multiple variables that contribute to susceptibility that span molecular to social science contributions. In Module 4 (Dose-Response Assessment) there is no reference to in vitro data (i.e. cell lines) – is this included? Perhaps this is what is referred to as high-throughput.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

I recognize the importance of the immune system in assessing toxicity and support the addition of immunotoxicity to the endpoint specific module group. However, sub-modules for immunotox and the other endpoints would be helpful to better define whether the focus here is on the mechanism of action or some other 'endpoint' (i.e. inflammation, autoimmune, infection susceptibility). Note that inflammation is a process that can lead to cancer or other endpoints, and these will have to be somewhat detangled. Also, endocrine seems missing as an endpoint.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

The mapped timeline and prioritization seems reasonable.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these

or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

I would prioritize low-dose extrapolation approaches.

Dr. Mara Seeley

Charge Questions: Human Toxicity Assessment Guideline

1. *Comment on EPA's proposed modular approach to developing their Consolidated Human Toxicity Assessment Guideline*

    - Using a modular approach to develop updated risk assessment guidelines makes a lot of sense, as it will facilitate timely updates of individual modules in conjunctions with advancements in specific aspects of risk assessment.
    - For peer review of certain modules (e.g., immunotoxicity) EPA should consider whether the SAB has the requisite technical expertise, or if ad-hoc committees/panels of relevant subject matter experts should be convened for the peer review.
    - In the 'Options for Module Development' text box in Figure 1, consider adding an additional bullet under item 2, for consultation with outside subject matter experts (e.g., from academia, NIEHS).

2. *Comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules"*

    - Overall, the common element modules are adequate, complete and well organized.
    - For Module 3, add that the systematic review conducted under Module 2 can inform Hazard Identification, with more weight given to higher quality studies.
    - Under Module 4, it is not clear how a margin of exposure would be derived without an evaluation of exposure, which doesn't seem to be a component of the common element modules.

3. *Comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's "endpoint-specific modules"*

    - Key endpoints, which are common critical effects for chemicals in EPA's IRIS database, are not included, e.g., hepatotoxicity and renal toxicity.

4. *What modules should EPA work on first and why, including commentary on extent of update needed for existing guidelines*

    - The Immunotoxicity module should be prioritized, given the complexity of the immune system and potential challenges in interpreting certain immunological findings with respect to apical effects.
    - Within the Common Element Modules, Module 4 should be worked on first, as the information included in this module seems like it would be most likely to advance the state-of-the art for conducting toxicity assessments.

5. *Comment on which issues should be considered higher priority*

- Methods to harmonize d/r evaluation for cancer/noncancer, and use of various dose-response modeling approaches would be higher priority (in that order).

Dr. Kimberly White

Charge questions for the SAB consultation on EPA activities to re-examine and consolidate EPA's Human Toxicity Assessment Guideline

1. **Question: EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.**

Answer: I am encouraged that EPA has taken into consideration the 2019 feedback from the SAB and the numerous public comments in order to develop a more thoughtful approach to updating existing human health toxicity related guidelines. The agency has indicated that it intends to complete the design of the Consolidated Guideline and prioritize the modules to be developed in December 2020 and then it will initiate the development of the modules in January 2021. The Agency indicates that the Consolidated Guideline will focus only on hazard characterization and dose-response assessment. However, the Agency should include information regarding any future plans for addressing exposure assessment or risk characterization and how the plans for this Consolidated Guideline will be used along with those other elements of the risk assessment process. I would also encourage the Agency to include a list of all the existing Agency guidance documents that will be revised, updated or incorporated as part of this Consolidated Guideline, and update figure 1 to include the opportunities for public comment and peer review (in addition to the SAB consultations) associated with each phase of the process.

2. Question: Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Answer: The proposed "common element modules" appear to be reasonable starting points for development of various aspects of the Consolidated Guidelines. Below are some suggested recommendations for consideration on some of the identified modules.

- Module 1 Planning and Scoping a Human Toxicity Assessment – While the module description includes concepts like "fit for purpose" and problem formulation it should also include discussion of the application of the Consolidated Guidelines for various program office use. The program offices currently may be performing elements of risk assessment for varying regulatory purposes and that information should be discussed in this module. This module should also discuss where there are currently differences in program office approaches, and how the Consolidated Guidelines will seek to provide a unified or singular Agency approach.

- Module 2 Identifying and Evaluating Toxicity Studies – Suggest this module be renamed "Identifying and Evaluating Scientific Data" and that it include three sub-categories or modules focused on animal toxicity data; epidemiology data; and mechanistic data. Each one of these modules should discuss the (1) literature search process associated with the identification of primary peer reviewed publications, peer reviewed reviews or meta-analysis of primary data, and grey literature and (2) the data quality assessment (e.g. quantitative or qualitative assessment) and how the data quality information will used for interpretation within and between data streams.

- Module 3 Hazard Identification – This module should include: case study examples of how data could/will be integrated across data streams for the purpose of hazard identification, including how to integrate positive, negative and null data points; examples of adverse outcome pathways that the Agency will consider relying on and the level of data and confidence for plausible mode of action frameworks with relevant case examples; and information describing the weight of evidence framework and how that will be used in the determination of hazard.

- Module 4 Dose-Response Assessment – In addition to the areas included, this module should include: a review and discussion of the application of uncertainty factors; and specific case study examples which delineate the type of information that would be needed to move away from linear default assumptions in lieu of alternative approaches (i.e. non-linear dose response assessment).

3. Question: Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

Answer: Modules 5 – 11 – Generally, these appear to be the appropriate endpoints for focus. The agency should consider focusing on endpoints identified in modules 5 – 10. The agency should include discussion or subcategories in Module 5. Developmental Toxicity related to maternal toxicity, mortality, structural abnormalities, alterations to growth and functional impairment. Module 6. Reproductive Toxicity should also include sub-categories for female fertility and male fertility toxicity endpoints. Additionally, the agency should also consider including an endpoint for systemic toxicity (e.g. liver, kidney) and separately solicit public and peer review input for other endpoints of focus.

4. Question: EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Answer: EPA has identified a number of modules for inclusion in the Consolidated Guideline. The agency should focus on development of Modules 1– 4 as they will provide the foundation for the overall process. For the endpoint specific modules, all of these items are important but if the agency is unable to do them in parallel, suggest the agency evaluate upcoming regulatory decisions where updated endpoint specific guidance would be most beneficial.

5. Question: EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority: a. Use of various dose-response modeling approaches (e.g., model averaging); b. Further consideration of the use of low-dose extrapolation approaches; c. Additional consideration of endogenous production of environmental contaminants; and d. Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

Answer: EPA has identified several issues above that would be important to address in the development of the Consolidated Guidelines. The use and application of dose-response modeling approaches and dose- response extrapolation including case study examples of how data can be used to inform the dose-response assessment should be priority areas of focus. Also, as an additional area of focus is understanding impacts of endogenous production in determining human health risk given that the agency may be currently evaluating substances that are produced endogenously.

Dr. Richard Williams

EPA is to be congratulated on producing the Consolidated Guidelines and, in particular, the Weight-of-Evidence (WoE) guidelines. However, EPA might consider changing the name to the "Weight of Good Evidence" as suggested below.

First Suggestion

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

In Module 3, EPA will consider "integrating/weighing evidence" and in Module 4, EPA will additionally consider "Data Quality Considerations." To integrate or weight evidence, it is first necessary to evaluate the quality of that evidence. EPA's Guidelines (Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency) are vague about how EPA will ensure that evidence is of sufficient quality.

Because of EPA's extensive use of epidemiology, it is essential that EPA address the quality of evidence these studies provide. Understandingly, this is an enormous challenge. For example, EPA has recently addressed formaldehyde. A search in Google Scholar of "formaldehyde" produced 1.9 million papers. EPA scientists and economists need guidance to quickly sort research papers for high quality.

There are scales of evidence that can be applied relatively quickly although they are not necessarily dispositive. Nevertheless, failure to use high quality epidemiological evidence may lead to regulations that are excessively costly or even harmful, particularly where there are risk/risk tradeoffs.

Epidemiology is in a crisis mode, but this is not a new complaint. Issues (not just confined to epidemiology) include pressure to get publications (for tenure, continued funding or promotions); political and personal biases of journal editors; and pressure for only publishing positive studies (excluding negative or no association) or newsworthy studies. All of these lead to poor quality science. In trying to replicate epidemiology papers, researchers have found failures to identify confounders, statistical weakness, data dredging (p-hacking), unwarranted focus on relative versus absolute risks and, confusing hazards with risk.

Some have argued that the problem is industry funding, but a recent study looked at 5,675 clinical nutrition, food safety, dietary patterns and dietary supplement scientific papers and found a surprising conclusion. "Industry funding is not consistently associated with producing research results that are considered 'biased' using the standard ROB (risk of bias) criteria" as compared to government-funded research.[1]

[1] Myers, E.F., et al., "Using risk of bias domains to identify opportunities for improvement in food- and nutrition-related research: An evaluation of research type and design, year of publication, and source of funding," PLOS|ONE, July 5, 2018.

Epidemiological studies in particular suffer from being difficult to reproduce, failure to identify confounders (making them hard to apply to the general population); and a focus on cases where a subpopulation is exposed to high dose levels (again, making it hard to map these results to the general population).[2] Examples of such cases are studies of the effect of ionizing radiation relying on evidence from radiation exposure post Hiroshima, Nagasaki, Chernobyl, and Fukushima, occupational radiation studies, and medical studies on highly exposed individuals.[3]

Studies have also shown that dietary and environmental studies are weaker than other scientific fields, perhaps because there are more media and public interest in these areas.[4] For example, in far too many cases in environmental science, causation is asserted when, at best, there is only correlation.

Many of the summary statements about poor science in recent years point to epidemiological research. In 2005, Stanford University professor John Ioannidis said he believed that, "It can be proven that most claimed research findings are false."[5] Four years later, the editor the New England Journal of Medicine wrote, "It is simply no longer possible to believe much of the clinical research that is published, or to rely on the judgment of trusted physicians or authoritative medical guidelines."[6] The editor of another highly rated medical journal, Lancet wrote, "The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness."[7]

None of this is to argue against the importance of epidemiology. To the contrary, it is important enough that the focus should be to identify high quality studies for use in risk assessments.

Recommendation: Prepare Guidelines on assessing the quality of epidemiological research used in risk assessments.

Second Suggestion

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- *Use of various dose-response modeling approaches (e.g., model averaging);*
- *Further consideration of the use of low-dose extrapolation approaches;*

---

[2] Taubes, G. "Epidemiology Faces Its Limits." Science 269, no. 5221 (Jul 14, 1995): 164-9.

[3] Calabrese, E. J., and M. K. O'Connor. "Estimating Risk of Low Radiation Doses - a Critical Review of the Beir Vii Report and Its Use of the Linear No-Threshold (LNT) Hypothesis." Radiat Res (Oct 20, 2014).

[4] Kabat, Geoffrey, *Getting Risk Right,* Columbia University Press, New York, NY, 2017, p. 24.

[5] Ioannidis, John P. A., "Why Most Published Research Findings Are False," PLOS|MEDICINE, August 30, 2005

[6] Angell M. Drug Companies & Doctors: A Story of Corruption. The New York Review of Books magazine. [Last accessed August 5, 2015]. Available from: http://www.nybooks.com/articles/archives/2009/jan/15/drug-companies-doctorsa-story-of-corruption/

[7] Horton, Richard, "Offline: What is medicine's 5 sigma?" Lancet, 385, April 11, 2015.

- *Additional consideration of endogenous production of environmental contaminants; and*
- *Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.*

The first two should be considered high priority.

Dose-Response Modeling

Depending on the specific law governing an EPA decision, safety assessments may be sufficient whereas others will require a risk assessment. Safety and risk assessments are *different*; a safety assessment should never be called, or confused with, a risk assessment. A safety assessment gives zero information about risk as there may be positive risk above or below any chosen "safe" level of exposure.

EPA's definition of what is safe is just one definition. For example, a recent paper in *Risk Analysis* showed that for chronic noncancer human health reference (safety) values, values used by EPA, Health Canada, US Agency for Toxic Substances and Disease Registry and RIVM in the Netherlands, produced different values 74% of the time for the same chemicals.[8]

Because of the limitations, a safety assessment is useful only as a screening tool, i.e., its primary function should be to allow a product on the market without further review as the level of risk is *likely* to be *de minimis.* As long as exposure is expected to be less than the reference dose (RfD) or other safety levels, it can be quickly approved. If the chemical fails a safety assessment, a risk assessment should follow. No safety analysis should be used to deny product access to the market.

If a product is a candidate for regulation (its use might be attenuated or excluded), then it is necessary to determine actual risk. For any risk decision for which a Regulatory Impact Analysis is needed, neither a safety analysis nor a conservative risk model is warranted.[9] Although decision makers may wish to know the risk for highly sensitive or highly exposed people, risk assessors also need to provide an unbiased central or expected value of risk. With this information, economists can compare benefits and costs for regulatory options. This value will also allow for risk/risk comparisons so that new products may be compared to existing ones on the market and unintended consequences can be evaluated.

Low Dose Extrapolation – Thresholds and Hormetic models vs. LNT

Consideration of the use of low dose extrapolation methods should also be high priority. Linear, no threshold dose-response models should not be the default. Either a threshold or hormetic model is more likely than an LNT and, when in doubt, a threshold model should be the default. Since 1977, the one-hit theory, i.e., the LNT, has dominated regulatory risk assessments in federal health and safety agencies. Modern molecular biology has challenged this theory over

[8] Holman, Elizabeth et. al, "Part I—Comparing Noncancer Chronic Human Health Reference Values: An Analysis of Science Policy Choices," *Risk Analysis,* 2016.

[9] See, for example, Williams, Richard A and Kimberly Thompson, "Integrated Analysis: Combining Risk and Economic Assessments While Preserving the Separation of Powers," *Risk Analysis,* Vol 24, No. 6, 2004.

the last several decades and found that DNA is not inert but actively defending itself at the subcellular, cellular, organ and whole-body levels against cancer.[10] Other studies have shown that multiple hits do not necessarily additively accumulate over time.[11]

According to EPA Administrator Douglas Costle, the One-Hit model was chosen due to its conservative nature, i.e., a bias toward overestimation of risk in the presence of uncertainty.[12] A later publication suggested that wide application of the LNT model in regulatory risk assessment was due in part to its attractiveness to regulators, namely: "it is easy to apply and […] it will generate an upper bound on the unknown, underlying cancer risk in most instances".[13] Because of this bias, when the LNT is used in a regulatory benefit analysis, the benefits will be biased upward and most policy makers will not be made aware of that fact.

Hormesis

It is widely understood that, for example, vitamins and minerals are toxic at high doses yet beneficial at low doses. The same is true for water. It is necessary for life at a certain dose rate yet, too much too quickly results in death (hyponatremia). This property, hormesis, has been extensively documented, including for radiation. Positive effects from low dose exposure have found evidence of beneficial effects related to fecundity, plant growth, cancer, heart disease, and inflammation.[14]

The hormesis model hypothesizes that exposure to low doses of stressors is protective (i.e., beneficial) and only becomes harmful at higher doses. Hormetic doses have been observed 4 to 5-fold below NOAELs for particular endpoints such that most risk models will not discover this J-shaped function. In regulations, hormetic models should be used primarily when policy makers are considering either not allowing a new product on the market or banning an existing product.

Recommendation: Replace the Linear No Threshold Model as the default dose/response model with a threshold model or, when appropriate, consider a hormetic model.

[10] Golden, B., J. Bus and E. Calabrese, "An examination of the linear no-threshold hypothesis of cancer risk assessment: Introduction to a series of reviews documenting the lack of biological plausibility of LNT," Chemico-Biological Interactions, 301 (1), Mar 2019, pp 2-5.

[11] Golden, B., J. Bus and E. Calabrese, "An examination of the linear no-threshold hypothesis of cancer risk assessment: Introduction to a series of reviews documenting the lack of biological plausibility of LNT," Chemico-Biological Interactions, 301 (1), Mar 2019, pp 2-5.

[12] U.S. Environmental Protection Agency. "Health Risk and Economic Impact Assessment of Suspected Carcinogens. Interim Procedures & Guidelines." 21402-05. Federal Register, 1976.

[13] OSTP (Office of Science and Technology Policy). (1985) Chemical carcinogens: review of the science and its associated principles. Federal Register 50:10372-10442.

[14] Calabrese, E.J., et. al., "Hormesis: A Highly Generalizable and Reproducible Phenomenon with Important Implications for Risk Assessment," *Risk Analysis,* 1999.

Enclosure C

Individual Comments from Members of the EPA Science Advisory Board Chemical Assessment Advisory Committee on EPA's Proposed Approach for Developing a Consolidated Human Toxicity Assessment Guideline

Dr. Richard Belzer C-1

Dr. Tiffany Bredfeldt C-16

Dr. Karen Chou C-19

Dr. Harvey Clewell C-21

Dr. David Hoel C-25

Dr. Michael Jayjock C-26

Dr. Wayne Landis C-27

Dr. Dennis Paustenbach C-33

Dr. Ted Simon C-36

Dr. Eric Smith C-38

Dr. Laura Vandenberg C-41

Dr. Richard Belzer

Comments on SAB/CAAC Review of Proposed *Consolidated Human Toxicity Assessment Guideline*[1]

1. **EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.**

A modular approach is sensible for establishing common rules and procedures with which all toxicity assessments conform. Endpoint-Specific Toxicity Guidelines should differ only with respect to differences in the science, not any of these common elements.

This change could have salutary effects throughout U.S. EPA's risk analysis ecosystem if it is faithfully implemented. For example, it would have beneficial spillover effects on the Agency's Quality System.[2] Also, it could reinvigorate the Agency's commitment to the information quality principles of transparency (through reproducibility), utility, integrity, and objectivity to which it committed 18 years ago.[3] And perhaps most usefully, it could improve the quality of Agency peer review (including SAB review)[4] so that scarce reviewer time is not wasted on reviewing common elements.[5]

The proposed Consolidated Guideline framework would not remedy two key weaknesses of the current toxicity guideline regime, however. First, toxicity assessments often are *de facto* regulatory standards. They are widely used as defaults for regulatory standard-setting; both U.S. EPA and the States rely on them to set enforceable standards. Were this not so, few if any toxicity assessments would be controversial and the Office of Research and development (ORD) could complete each of them on schedule.

There surely are scientific controversies over which scientists are quite prepared to wage war – the Thirty (Sixty?) Years War over genotoxic and nongenotoxic modes of action in carcinogenicity comes to mind. But these intramural squabbles are not what drives public controversy. Rather, current toxicity guidelines are controversial because they contain substantial, if not controlling, risk management policy defaults. It will take forever for ORD to complete each module of the Consolidated Guideline project if the Office is determined to retain these features. The endeavor would be mortally wounded if the main debate concerns

[1] The charge to the committee is presented in U.S. EPA Science Advisory Board (2020).
[2] U.S. Environmental Protection Agency (2020a).
[3] Office of Management and Budget (2002), U.S. Environmental Protection Agency (2002).
[4] U.S. Environmental Protection Agency (2015).
[5] Of course, a review panel must ensure that a chemical-specific toxicity assessment actually complies with the common elements in Modules 1-4. Scarce peer reviewer resources would not be saved if this does not happen. To avoid this potential problem, ORD could organize its peer reviews of individual toxicity assessments in multiple parts, with one panel reviewing compliance with Modules 1-4 and other panels reviewing compliance with Each applicable endpoint-specific guideline.

whether precautionary risk *management* should continue to be embedded in ostensibly scientific risk *assessment*.

Second, ORD toxicity assessments are incompatible with regulatory benefit-cost analysis. To be valid, a benefit-cost analysis must provide objective characterizations of both baseline human health risk and the benefits and costs of alternative regulatory interventions.[6] As a matter of Agency policy, however, U.S. EPA toxicity assessments,[7] cancer risk assessments,[8] exposure assessments,[9] risk characterizations,[10] uncertainty analysis,[11] and risk assessment generally,[12] all purposefully overstate the expected value of risk.[13]

Whether these policies have merit for risk *management* is immaterial; they have no merit whatsoever for regulatory benefit-cost analysis. Acknowledging this historic and pervasive bias may be perceived as opening up Pandora's Box, but let's be clear. Not only has the Agency saturated its ostensibly scientific practices with nonscientific policy choices, it also has scrupulously maintained the strictest possible separation between risk assessment and economics. Only in rare cases (if ever) have the outputs of ORD toxicity assessments been useful for estimating either the social costs of baseline risk or the expected benefits of alternative regulatory interventions.

---

[6] Unbiasedness has been widely advocated by professional economists (see, e.g. Arrow et al. 1996; Dudley et al. 2017); prescribed in government-wide guidance (Office of Management and Budget 1988, 1990a, 1990b, 1996, 2000, 2003); and incorporated within USEPA guidance (U.S. Environmental Protection Agency 1983, 2000, 2010, 2014, 2016).

[7] U.S. EPA Office of the Science Advisor (2004, p. 53): "Unless there are data to indicate otherwise, a change that is considered adverse (i.e., associated with toxicity) is assumed to indicate a problem for humans."

[8] U.S. EPA Office of the Science Advisor (2004, p. 45): "An evaluation should be made as to whether low-dose linear extrapolation is sufficient to protect susceptible populations."

[9] U.S. EPA Office of the Science Advisor (2004, Sec. 2.2.2 ["Whom Is EPA Trying to Protect?"] and Sec. 2.2.7 ["How Are High-End Exposures Reflected in EPA Evaluation?"]).

[10] U.S. EPA Office of the Science Advisor (2004, p. 10): "Our Risk Characterization Policy directs us to consider all scientifically *plausible* and *supportable* viewpoints" (emphasis added). See also p. 13: "Many comments to EPA suggest that the combining of upper ends leads to unreasonable estimates of risk. We generally believe otherwise…" and U.S. Environmental Protection Agency Science Policy Council (2000, Sec. 1.3.4), which substitutes *reasonableness* (a subjective term) for *unbiasedness* (which has an objective definition).

[11] U.S. EPA Office of the Science Advisor (2004, Sec. 4.6.4).

[12] U.S. EPA Office of the Science Advisor (2004, p. 13): "[S]ince EPA is a health and environmental protective agency, EPA's policy is that risk assessments should not knowingly underestimate or grossly overestimate risks. This policy position prompts risk assessments to take a more 'protective' stance given the underlying uncertainty with the risk estimates generated. Another framing policy position is that EPA will examine and report on the upper end of a range of risks or exposures when we are not very certain about where the particular risk lies… [W]hen several parameters are assessed, upper-end values and/or central tendency values are generally combined to generate a risk estimate that falls within the higher end of the population risk range."

[13] U.S. EPA Office of the Science Advisor (2004, p. 13): "[D]efault assumptions utilized in any given risk assessment entail science policy positions or choices. These science policy choices are more specific than the framing science policies, but generally are consistent with the framing policies. For example, a change that is considered adverse (i.e., associated with toxicity) in an animal study is assumed to indicate a problem for humans unless data demonstrate otherwise."

The result has been a systematic inability to use toxicity assessments to evaluate the consequences of alternative policy choices. The proposed Consolidated Guideline framework displays little or no awareness of this problem, or what would be much worse, an unstated desire not to solve it. ORD staff have made it clear that an extraordinary amount of work will be required to complete this project. But that work will be in vain if toxicity assessment is not – finally – integrated with benefits assessment. An approach that translates scientific knowledge directly into benefit estimates could be a way to accomplish this, for at least that way the consequences of *de facto* regulation would be transparent, but that is not the direction implied by the proposed Consolidated Guideline.

For this reason, I strongly urge ORD to ensure that Module 1 includes the biological prerequisites for benefits assessment, and that similar changes are made in subsequent Modules. Every Endpoint Specific Module also must be so informed, of course, but the most demanding, challenge will be integrating economic principles into Module 4 ("Dose-Response Assessment"). This will require a body of expertise that ORD does not have. One of our tasks will be to recommend ways for the Office to gain and apply it.

2. Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

A. Modules 1-4 (but especially Module 1)

Proposed Module 1 ("Planning and Scoping a Human Toxicity Assessment") appears to include most of the expected "key elements." The Agency should clearly define all of these "key elements," and adherence to these definitions must be both required and objectively refutable. Obviously, "key elements" that are not objectively defined, or are subjectively interpreted, would destroy the uniformity across common elements that a Consolidated Human Toxicity Guideline requires to be successful.

Some "key concepts" identified in Module 1 (e.g., "cumulative risk") may be difficult to objectively define. Historically, cumulative risk has been constrained to risks sharing a common mechanism of toxicity.[14] This scope is inherently incomplete several ways. It systematically excludes indirect risks to human health and welfare resulting from action to reduce a chemical risk (i.e., "risk-risk" and "health-health" tradeoffs)[15] and substitution risks.[16]

[14] U.S. Environmental Protection Agency (2003). More generally, *cumulative risk* means "the combined risks from aggregate exposures to multiple agents or stressors" (p. xvii), but as noted below, in practice it tends to be Gerrymandered to exclude a host of risk-related tradeoffs.

[15] See, e.g., Keeney (1990, 1994), Lutter and Morrall III (1994), Viscusi (1994a), and Viscusi (1994b).

[16] See, e.g., Viscusi (1988, pp. 69-83), Viscusi (1994b), and (less technically) Graham and Wiener (1995) and Viscusi (2018, pp. 133-136).

The rule in benefit-cost analysis is that every benefit and every cost should be counted, with each benefit and cost counted exactly once. This goal is unachievable in practice because not all benefits and costs can be quantified, and not all quantified benefits and costs can be monetized. But technical limitations do not justify abandoning the rule. Rather, they argue for concerted effort to better identify, quantify, and monetize what's missing. It is unhelpful to devote resources toward adding yet another significant figure in the estimation of a well-understood benefit or cost while important benefits and costs remain unquantified or unmonetized. Economists may be especially familiar with the so-called "drunk and lamp post problem," but it applies well to risk analysis:

> A drunk loses his keys and is looking for them under a lamp post. A policeman comes over and asks what he's doing.
>
> "I'm looking for my keys," he says.
>
> "Where did you lose them?" the policeman asks.
>
> "I lost them over there."
>
> The policeman looks puzzled. "Then why are you looking for them over here?"
>
> "Because the light is so much better here."[17]

U.S. EPA should resist the temptation to look for its "keys" underneath the lamp post. Rather, the Agency should be guided by a rigorous evaluation of the value of information. Which has more net social value: (1) marginal improvements in an existing endpoint-specific module, or (2) creating a module where none currently exists? For an existing endpoint-specific module, which has more net social value: (1a) making marginal improvements within the existing structure, or (1b) overcoming the deadweight loss that has accumulated over decades of relentless drift in upward bias, excess precision, absent or understated characterization of uncertainty, and unsupported causality assumptions? It's been said that "success consists of going from failure to failure without loss of enthusiasm."[18] This is not a healthy path for risk assessment.

Some "key concepts" listed in Module 1 (e.g., "vulnerable populations") may not be capable of objective definition. The glossary in U.S. Environmental Protection Agency (2019) borrows definitions of "vulnerable population" from NLM, the Centers for Disease Control, and the Resilience and Adaptation in New England (RAINE) Glossary. Each is subjective. Indeed, any difference in toxicity within the population could be interpreted as a manifestation of lesser or greater "vulnerability." However, because the typical purpose of identifying "vulnerable populations" is to give them special (i.e., subjectively higher) policy weight, it is hard to imagine how this concept could ever be defined objectively. Module 1 should not include any purportedly "key elements" that are subjectively defined. Subjectivity in Module 1 invites subjectivity in

[17] A representative version of the joke is related by Leaver (2014).

[18] Freedman (2010a), excerpted at Freedman (2010b), who attributes the aphorism to Winston Churchill.

every subsequent module. If subjective "key elements" are included, the Common Element Module will not be scientific.

As noted in my response to Question 1, one of the "key elements" listed in Proposed Module 1 is "fit for purpose" (elsewhere "fitness for purpose").[19] A key purpose of risk assessment is the estimation of benefits for priority-setting, regulatory standard-setting, and similar activities. But ORD toxicity assessment is generally not fit for these purposes. As noted above, toxicity assessment is incompatible with benefit-cost analysis.

Lest this concern appear parochial to benefit-cost analysis, it should matter to ORD even if the Office is determinedly opposed to making its outputs strictly scientific. To make rational choices among alternative ways to expend scarce resources on the development of chemical-specific toxicity guidelines, ORD needs a way to rank alternatives. A logical way to rank them is in declining order of net social benefits. This cannot be done for two key reasons: (1) risk estimates are upwardly biased by variable but unknown amounts, and (2) ORD has no clue about net social benefits even if risks were ranked without bias. Overcoming both requires economic reasoning, and in its absence ORD is compelled to set priorities based on internal or external politics.[20] Thus, Module 1 of the Consolidated Human Toxicity Guideline must include provisions sufficient to ensure that all modules strive for (and not eschew) unbiasedness.[21]

Adherence to information quality principles[22] also is not included in the list of "key elements." It should be. These principles apply to all influential information disseminated by U.S. EPA, and it should go without saying that ORD toxicity assessments are "influential."[23] In addition, every endpoint-specific module will have information quality concerns, and applicable information quality guidelines are neutral with respect to all of them. U.S. EPA must ensure that every module adheres to these guidelines and includes effective procedures for pre-dissemination review to prevent error and procedures to correct errors that nevertheless occur.

Data quality is mentioned in proposed Module 4 ("Dose-Response Assessment"), but this is likely to be too late and too selective. The quality issues related to toxicological data are not unique; they exist in Module 2 ("Identifying and Evaluating Toxicity Studies"), Module 3 ("Hazard Identification"), and in every proposed endpoint-specific module. It would be much better to incorporate information quality concerns in Module 1 so that all subsequent modules (and all implementations of them) are treated the same.

Finally, a key element absent from Proposed Module 1 is humility. Whether by self-selection,

[19] Office of Management and Budget (2019), U.S. Environmental Protection Agency (2020b).

[20] Much of the June 23-24 public meeting was devoted to answering Charge Question 4, which concerns how ORD should set priorities. Several competing proposals were made, and each was supported solely by the political, personal, professional, or institutional interests of the proposer – not the best interests of the public.

[21] Expected values are required whenever the entire risk distribution is not objectively characterized.

[22] Office of Management and Budget (2002), U.S. Environmental Protection Agency(2002).

[23] Office of Management and Budget (2002, p. 8460): "'Influential', when used in the phrase 'influential scientific, financial, or statistical information', means that the agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions."

training, or experience, many scientists (and perhaps especially risk assessors) suffer deficiencies of this quality. Risk estimates – remarkably, including low dose extrapolations orders of magnitude outside the boundaries of available data and across species, where scientific uncertainty and the temptations of mathematical delusion are the greatest – are routinely reported as if they are reliable, if not actually true. More than three decades ago, U.S. EPA sensibly characterized low-dose cancer risk estimates with the caveat that the true risk could be as low as zero.[24] The Agency abandoned this without scientific justification and adopted a policy preference favoring hubris.[25]

It's worth discussing how to imbue Module 1 with a spirit of humility, for there is no obvious mechanism or internal regulatory procedure through which it can be ensured. Nonetheless, humility is likely to be a genuinely "key element." Absent humility about the limits of scientific knowledge and the boundary between science and policy, ORD toxicity assessments will continue to be plagued by controversy, conflict, and limited productivity – no matter how (or even if) the Agency implements its proposed Consolidated Human Toxicity Guidelines.

B. Module 2

Module 2 concerns only collecting the database of toxicity studies from which ORD will conduct hazard identification (Module 3) and dose-response assessment (Module 4). Nonetheless, the practice of identifying and evaluating toxicity studies has been a source of sustained controversy. The summary in Table 1 suggests that nothing material is expected to change, even if the text of the module grows by leaps and bounds.

Module 2 must include rigorous compliance with applicable information quality guidelines. Those guidelines begin with transparency, which is measured by the performance standard of reproducibility.[26] Full access to data, models, and code is required to make reproducibility possible. Toxicity studies for which access is not available should not be included.[27]

C. Modules 3 and 4

As the text in Table 1 suggests, a considerable amount of territory is included in hazard identification, and much of it will be susceptible to conflict resulting from the application of policy judgment. "Weighing evidence" and "synthesizing results" can be done myriad ways. I strongly urge ORD to establish principles that discourage this to the greatest extent possible. A

[24] U.S. Environmental Protection Agency (1986). This caveat was abandoned when cancer risk assessments were incorporated as inputs to Agency benefit assessments.
[25] U.S. Environmental Protection Agency (1996, 2005).
[26] (Office of Management and Budget 2002, p. 8460): "'Reproducibility' means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased)… With respect to analytic results, 'capable of being substantially reproduced' means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error."
[27] Special access provisions are reasonable for data that constitute trade secrets or otherwise qualify as confidential business information. For more on this subject, see Belzer (2020).

policy of intellectual neutrality should apply, and procedures should be created to identify departures from that norm, and correct them.

Dose-response assessment faces similar, and probably much greater, challenges. Several items in the list of issues are scientific, but others are not. Specifically, how to perform benchmark dose modeling and what to do with RfDs (which are economically uninterpretable) will be difficult. The changes I have proposed for Module 1 could go a long way toward making Modules 3 and 4 less fraught with conflict, but that requires ORD to first decide that toxicity assessments will henceforth be compatible with benefit-cost analysis.

3. **Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.**

The choice or sequencing of endpoint specific modules (Modules 5–11 in Table 1) should be made based on value-of-information (VOI) principles, which are set forth as statutory elements of Federal information policy.[28] These principles should be familiar to ORD because virtually every request or demand for information from the public is subject to applicable statutory and regulatory requirements. The Office has five active OMB Control Numbers.[29]

Discussion during the June 23-24, 2020, public meeting was not informed by these statutory requirements. Nonetheless, VOI principles should guide ORD's evaluation of existing modules, the potential identification of additional modules, and the sequencing of their creation or revision. The general principle is straightforward: the practical utility of a new or revised endpoint-specific module must exceed its burden, where *practical utility* in this case means the expected public health benefits of the module, and *burden* means its opportunity costs. Modules (or components thereof) would be ranked by their net practical utility, and ORD would sequence the development of modules in order of declining net practical utility.

To be clear, a ranking in order of net practical utility is not the same as a ranking that would be obtained by giving priority to endpoints for which no guidelines currently exist. The absence of a guideline does not imply practical utility. Nor would a net practical utility ranking be the same as a ranking that would be obtained by asking risk assessors to rank their preferences. The interests of risk assessors can be highly parochial. For example, cancer risk assessors might reasonably believe that revisions to Module 8 ("Carcinogenicity") are most important, whereas neurotoxicity risk assessors could think that Module 10 ("Neurotoxicity") belongs at the top. Choosing among these modules based on the relative strength of preference among risk assessors makes the outcome dependent on which risk assessors are polled and who does the polling. If instead endpoint-specific modules are ranked and selected based on their net practical utility, their value to the public (which, let us remember, is supposed to guide Agency decision-making) can be taken into account.

[28] Paperwork Reduction Act of 1995 (44 U.S. § 3501 *et seq*.), implemented via the Information Collection Rule (5 U.S.C Part 1320).

[29] 2080-0005, 2080-0021, 2080-0082, 2080-0083, and 2080-0084.

For any endpoint-specific module to produce public health benefits, it must generate outputs capable of conversion into appropriate public health units. This means, as a threshold matter, that an endpoint must be comprehensible to the public and susceptible to valuation. Endpoints that nonexperts do not understand are difficult or impossible to value. Thus, a module that addresses an endpoint that nonexperts cannot understand or value has little or no practical utility.

To be clear, a lack of public understanding does not condemn an endpoint to irrelevance. Rather, it draws attention to research needs that may be simultaneously hidden and urgent. Suppose risk assessors can agree that a particular endpoint is crucial for estimating a human health risk, but currently it is not comprehensible to the public and thus not susceptible to valuation. To aid rational decision-making, more must be learned (and quickly) to overcome these knowledge gaps. Only then can Agency decision-makers properly elevate this endpoint to the stature it deserves in the risk management agenda.[30]

4. EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

As noted above in my response to Question 3, a *scientific* way to set priorities among endpoint-specific modules is based on VOI principles. That means maximizing net practical utility. Note that practical utility in this context is just a synonym for expected public health benefits. To allocate scarce Agency resources any other way means achieving less protection of public health.

Suggestions regarding near-term priority-setting are provided at the end of these comments in the section titled "Recommendations."

5. EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

A. Use of various dose-response modeling approaches (e.g., model averaging)

Model averaging can be a valuable way to reduce bias, but it requires that the models being "averaged" (i.e., weighted) be independent and not strategically biased. If they are not independent, then their biases are correlated; and if their biases are correlated, averaging them may increase total bias. Averaging models that are strategically biased rewards bias rather than reducing it.

Model averaging (i.e., weighting) almost certainly won't be done objectively; someone will choose the weights. Should it be a risk assessor or a policy official? If risk assessors are

[30] It is certainly possible to use political pressure to elevate a publicly incomprehensible endpoint to the top of the regulatory agenda. However, risk assessors should be wary of employing such tactics lest they lose their credibility as scientists.

scientists, they have no business exercising policy judgment (even if they'd like to do so). If it is a matter of policy judgment, then the authority and responsibility belongs to duly appointed Agency officials (even if they'd prefer to "follow the science" so they won't be held accountable for their policy choices).

My sense is that a better approach, and one less fraught with peril to the risk analysis profession in general and ORD toxicity assessments in particular, is to report all available models along with the available evidence for and against each one. Similar reporting schemes elsewhere have faltered because of bias that results when the party preparing the evidence is not (or is not perceived as) neutral. The information quality standards of presentational and substantive objectivity are handy here.[31] Substantive objectivity is achieved when information is "accurate, reliable, and unbiased." Presentational objectivity requires that information be "presented in an accurate, clear, complete, and unbiased manner." The latter includes proper context, for context easily goes by the wayside even when information is substantively objective. A way to reduce bias in presentation is to reform U.S. EPA peer review practices so that (fully reported) conflicts of interest are expressly encouraged, with the objective being to secure agreement among competing interests as to how evidence is presented.[32]

B. Further consideration of the use of low-dose extrapolation approaches

Low-dose extrapolation (more generally, extrapolation outside the bounds of available data) is always scientifically perilous. It has been done so often and for so long in human health risk assessment that we are desensitized to the intellectual peril. Recent experience with SARS-CoV-2 ("COVID-19") has shown just how damaging such extrapolation can be. On March 16, 2020, the Imperial College London (ICL) COVID-19 Response Team predicted that an "unmitigated epidemic" "would result in 2.2 million U.S. fatalities and 510,000 U.K fatalities, "not accounting for the potential negative effects of health systems being overwhelmed on mortality." Almost all of these fatalities would occur before August 20, with a peak daily death rate of about 17 per 100,000 that was forecast to occur about June 20.[33]

The data tell a very different, and much less dramatic, story. As of July 29, approximately 150,000 U.S. deaths associated with COVID-19 have been reported (6.8% of the ICL forecast).[34]

---

[31] Office of Management and Budget (2002, Sec. V(3)).

[32] The conventional peer review model (see, e.g., The National Academies 2003; Office of Management and Budget 2005), and U.S. EPA's Peer Review Handbook (U.S. Environmental Protection Agency 2015) treat conflicts of interest as liabilities rather than assets. And they are liabilities if the purpose of government peer review is to ratify agency preferences and decisions. If instead the purpose of peer review is to conduct peer review, conflicts of interests are assets. That's because those with conflicts of interest tend to be the most motivated peer reviewers. As long as they are not anonymous, they can responsibly hold their intellectual "foes" to the most rigorous scientific standards. When each "side" does this to the other, the quality of everyone's science improves.

[33] Ferguson et al. (2020, p. 7 [Figure 1]). They also estimated 1.1–1.2 million U.S. fatalities "even if all patients were able to be treated" (p. 16).

[34] Johns Hopkins University Coronavirus Resource Center (2020; "Cumulative Cases," accessed July 29, 2020). It's likely that some deaths caused by COVID-19 are not included. It's also certain that many deaths *attributed to* COVID-19 were not *caused by* COVID-19. Equally important, "cases" (i.e., positive tests) are problems only insofar as they have substantial associated health costs (e.g., hospitalizations), and may be benefits if it turns out they result in herd immunity.

Meanwhile, it has been estimated that through July 26 COVID-19 has resulted about in $310 billion in mortality costs while the governmental responses have imposed $2.1 trillion in costs – a benefit-cost ratio of roughly 15%.[35]

Experience with COVID-19 is different from chemical toxicity assessment for many reasons, but the reason most relevant here is that the predictions ICL made beyond the boundary of the data were testable after the fact, and U.S. EPA low-dose risk assessments generally are not. Based on this experience, however, the Agency should reconsider how much confidence in low-dose extrapolation is scientifically justified. It also should do more to accurately characterize uncertainty on key margins, including causality, and develop a practice of humility with respect to scientific uncertainty that is evidenced by full disclosure.

C. Additional consideration of endogenous production of environmental contaminants

The endogenous production of toxic substances (e.g., formaldehyde) reduces the expected value of baseline risk. Failing to account for this inflates risk reductions from lowered environmental exposure. If ORD adopted methods that were compatible with benefit-cost analysis, this problem would go away. Risks from endogenous production would be accounted for in the baseline, and not attributed to environmental exposure (unless the public were willing to pay to avoid the endogenous production of environmental toxicants).

D. Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

There is no justification for making any conceptual distinction between cancer and noncancer dose-response. A biological phenomenon is adverse if and only if an optimally informed person is willing to pay to avoid it. The nature of a risk is relevant only insofar as it affects a health endpoint. The severity of a risk is fully captured by the magnitude of willingness-to-pay (WTP).[36] Thus, the purpose of risk assessment is to estimate *first* the loss imposed by a risk on human welfare, and estimate *second* the welfare gain expected to be realized by reducing it. The purpose is never to derive a "worst case" (or some variant thereof) for the baseline risk, a "best case" for the amount of risk reduction that would be achieved by intervention, or the divination of what exposure is "safe." These objectives may have scientific inputs, but they are inherently nonscientific activities.[37]

A major problem remains because WTP depends on the quality of lay risk comprehension. Some risks (e.g., premature mortality, financial harm) are well understood by nonexperts, but many are not. There surely are phenomena (some biological) that scientists and risk assessors are able to understand sufficiently well to comprehend them as adverse. But for a republican

[35] Mulligan (2020; accessed July 29, 2020).

[36] Willingness-to-pay (WTP) is the maximum price at which a person will voluntarily engage in an exchange. This is the foundation of welfare economics. Benefit-cost analysis (and U.S. EPA's practice of Regulatory Impact Analysis) is built on this concept.

[37] "Safety" has no scientific definition. It is inherently controlled by policy preferences, on which members of the public hold diverse but equally legitimate views. The entire safety assessment edifice is unsustainable as scientific risk assessment.

government in a democratic society, the authority for making risk-reduction decisions cannot be delegated or usurped by a scientific clerisy. It is our job as scientists and risk assessors to develop ways to translate complex and presumably risky phenomena into popularly understandable forms that enable nonexperts to credibly value risk reduction or prevention. And we must do so without embedding *our* risk preferences along the way.[38]

With this in mind, U.S. EPA should focus the development or revision of endpoint-specific modules where nonexperts already have sufficient knowledge and insight to value risk reduction. The Agency should postpone the development of modules where this knowledge is lacking until technologies have been developed, tested, and validated that effectively and objectively translate complex endpoints into language nonexperts can comprehend. To expedite the transformation of modules from the second to the first group, a crash course of research may be warranted to develop these transformational technologies.

6. Recommendations

The completion of a Consolidated Guideline might take years, as ORD staff and some public commenters have predicted. But completion should not be ORD's current focus. Rather, the Office should devote all its attention now to getting the main moving parts of Module 1 done. These moving parts are:

A. Defining key concepts, with the explicit inclusion of scientific objectivity as the overarching principle to which every other concept and element of the Consolidated Guideline must be subordinate.

B. Ensuring that the Guideline outputs are fully compatible with benefit-cost analysis, and thus can be used to objectively estimate and rank baseline risks, and objectively estimate the expected benefits of alternative changes in actual human exposure.

C. Adhering to information quality principles and standards.

D. Embracing VOI principles (like those established by the PRA) as the basis for ranking the creation or revision of endpoint-specific modules.

E. Conducting or sponsoring the research needed to enable endpoints to be understood and valued by nonexperts so that the development of guidelines for them has practical utility.

Committing the first four of these principles to internally binding policy can and should be completed before the end of 2020, and the fifth should be made a research priority for FY 2022 budgeting. ORD should not be so distracted by the magnitude of the entire endeavor that it fails to construct the foundation on which a Consolidated Guideline must be built.

[38] COVID-19 offers lessons here as well. Public health experts have caused enormous damage to their reputations by professing knowledge instead of admitting ignorance, advising policies and actions based on speculation or personal policy preferences but attributing it to Science, and displaying rampant personal and professional hypocrisy. A substantial fraction of the public appears to have tuned them out.

References

Arrow, Kenneth J., Maureen L. Cropper, George C. Eads, Robert W. Hahn, Lester B. Lave, Roger G. Noll, Paul Portney, Milton Russell, Richard Schmalensee, and Kerry Smith. 1996. "Benefit-cost Analysis in Environmental, Health, and Safety Regulation," American Enterprise Institute: Washington, DC.

Belzer, Richard B. 2020. "Comments on *Strengthening Transparency in Regulatory Science; Supplemental Notice of Proposed Rulemaking*." Richard B. Belzer: Mount Vernon VA. Available: https://www.regulations.gov/document?D=EPA-HQ-OA-2018-0259-11381.

Dudley, Susan, Richard Belzer, Glenn Blomquist, Timothy Brennan, Christopher Carrigan, Joseph Cordes, Louis A. Cox, Arthur Fraas, John Graham, George Gray, James Hammitt, Kerry Krutilla, Peter Linquiti, Randall Lutter, Brian Mannix, Stuart Shapiro, Anne Smith, W. Kip Viscusi, and Richard Zerbe. 2017. "Consumer's Guide to Regulatory Impact Analysis: Ten Tips for Being an Informed Policymaker," *Journal of Benefit-Cost Analysis*: 1-18. Available: https://www.cambridge.org/core/journals/journal-of-benefit-costanalysis/article/consumers-guide-to-regulatory-impact-analysis-ten-tips-for-beingan-informed-policymaker/FAF984595B822A70495621AEA7EF7DEB/core-reader.

Ferguson, Neil M, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, ArranHamlet, Wes Hinsley, Lucy C Okell, Sabine van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick GT Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. 2020. "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand," Imperial College COVID-19 Response Team: London UK. Available: https://www.imperial.ac.uk/mrc-globalinfectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/.

Freedman, David H. 2010a. *Wrong: Why Experts Keep Failing Us -- and How to Know When Not to Trust Them*. Hatchette Book Group of Little, Brown: New York NY.

———. 2010b. "'Wrong'." *New York Times,* June 10. Graham, John D., and Jonathan Baert Wiener (eds.). 1995. *Risk vs. Risk.* Harvard University Press: Cambridge MA.

Johns Hopkins University Coronavirus Resource Center. 2020. "Cumulative Cases." JHU: Baltimore MD, Accessed July 29. https://coronavirus.jhu.edu/data/cumulative-cases.

Keeney, Ralph L. 1990. "Mortality Risks Induced by Economic Expenditures," *Risk Analysis*, 10: 147-59.

———. 1994. “Mortality Risks Induced by the Costs of Regulations,” *Journal of Risk and Uncertainty*, 8: 95-110.

Leaver, Sean. 2014. "The Drunk Neo-classical Economist and the Lamp Post." *The misBehaving Economist*. Available: https://themisbehavingeconomist.com/2014/01/03/the-drunk-neo-classicaleconomist-and-the-lamp-post/.

Lutter, Randall, and John F Morrall III. 1994. “Health-Health Analysis: A New Way to Evaluate Health and Safety Regulation,” *Journal of Risk and Uncertainty*, 8: 43-66.

Mulligan, Casey. 2020. “U.S. Daily Cumulative Costs of the COVID-19 Pandemic, Updated Through July 26, 2020,” Accessed July 29, 2020. http://pandemiccosts.com.

Office of Management and Budget. 1988. “The Role of Economic Analysis in the Development of Responsible Regulation.” In *Regulatory Program of the United States Government, April 1, 1988 -- March 31, 1989.* Office of Management and Budget: Washington DC.

———. 1990a. “Regulatory Impact Analysis Guidance (Appendix V).” In *Regulatory Program of the United States Government, April 1, 1990 -- March 31, 1991*. Office of Management and Budget: Washington DC.

———. 1990b. “Regulatory Impact Analysis Guidance: Discussion of Comments.” In *Regulatory Program of the United States, April 1, 1990 -- March 31, 1991*. Office of Management and Budget: Washington, DC.

———. 1996. “Economic Analysis of Federal Regulation Under Executive Order 12866,” Accessed July 6, 2020. https://obamawhitehouse.archives.gov/omb/inforeg_riaguide.

———. 2000. “Guidelines to Standardize Measures of Costs and Benefits and the Format of Accounting Statements.” Office of Management and Budget: Washington, DC.

———. 2002. “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication,” *Federal Register*, 67: 8452-60.

———. 2003. “Circular A-4: Regulatory Analysis,” Office of Management and Budget: Washington DC, Accessed July 6, 2020. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A4/a-4.pdf.

———. 2005. “Final Information Quality Bulletin for Peer Review,” *Federal Register*, 70: 2664-67.

———. 2019. "Improving Implementation of the Information Quality Act." *M-19-15*. Office of Management and Budget: Washington DC.

The National Academies. 2003. "Policy on Committee Composition and Balance and Conflict of Interest." National Academies Press: Washington DC.

U.S. Environmental Protection Agency. 1983. "Guidelines for Performing Regulatory Impact Analysis." U.S. EPA: Washington DC.

———. 1986. "Guidelines for Carcinogen Risk Assessment," *Federal Register*, 51: 33992-4003.

———. 1996. "Proposed Guidelines for Carcinogen Risk Assessment," *Federal Register*, 61:17960-8011.

———. 2000. "Guidelines for Preparing Economic Analyses." In *EPA 240-R-00-003*. U.S. EPA: Washington, DC.

———. 2002. "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency " *EPA/260R-02-008*. U.S. EPA: Washington DC.

———. 2003. "Framework for Cumulative Risk Assessment." *EPA/630/P-02/001F*. U.S. EPA/Risk Assessment Forum: Washington DC.

———. 2005. "Guidelines for Carcinogen Risk Assessment." *EPA/630/P-03/001F*. U.S. EPA/Risk Assessment Forum: Washington DC.

———. 2010. "Guidelines for Preparing Economic Analyses." *EPA 240-R-10-001*. U.S. EPA/National Center for Environmental Economics: Washington DC.

———. 2014. "Guidelines for Preparing Economic Analyses." U.S. EPA/National Center for Environmental Economics: Washington DC.

———. 2015. "Peer Review Handbook." In *EPA/100/B-15/001*. Washington DC: U.S. EPA/Science and Technology Policy Council.

———. 2016. "Guidelines for Preparing Economic Analyses." U.S. EPA/National Center for Environmental Economics: Washington DC.

———. 2019. "Terms & Acronyms," U.S. Environmental Protection Agency: Washington DC. Accessed June 16, 2020. https://iaspub.epa.gov/sor_internet/registry/termreg/searchandretrieve/termsandacronyms/search.do.

———. 2020a. "Agency-wide Quality System Documents." U.S. EPA: Washington DC. Accessed June 16, 2020. https://www.epa.gov/quality/agency-wide-qualitysystem-documents.

———. 2020b. "Guidelines for Preparing Economic Analyses (SAB Review Draft)." U.S. EPA/National Center for Environmental Economics: Washington DC.

U.S. Environmental Protection Agency Science Policy Council. 2000. "Risk Characterization Handbook." *EPA 100-B-00-002*. U.S. EPA: Washington DC.

U.S. EPA Office of the Science Advisor. 2004. "An Examination of EPA Risk Assessment Principles and Practices; Staff Paper " *EPA/100/B-04/001*. U.S. EPA/Risk Assessment Task Force: Washington DC.

U.S. EPA Science Advisory Board. 2020. "Charge to the SAB on the Consolidated Human Toxicity Assessment Guideline." U.S. EPA/SAB: Washington DC.

Viscusi, W. Kip. 1988. *Rational Risk Policy.* Clarendon Press: Oxford.

———. 1994a. "Mortality Effects of Regulatory Costs and Policy Evaluation Criteria," *The RAND Journal of Economics*: 94-109.

———. 1994b. "Risk-Risk Analysis," *Journal of Risk and Uncertainty*, 8: 5-18.

———. 2018. *Pricing Lives.* Princeton University Press: Princeton NJ.

Dr. Tiffany Bredfeldt

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

a. *The choice to consolidate guidelines is one for which EPA can be applauded. This should improve efficiency and transparency of approaches. The decision to utilize a modular approach to the consolidated guidelines is also favorable as it should be a more focused and efficient way to tackle the challenge of consolidating the guidelines.*

b. *The proposed approach is reasonable and appears to be a very logical path forward. The process as shown in Figure 1 is also very logical, particularly in prioritization of modules to be developed first. The timeline appears to be ambitious, but I do believe it to be achievable.*

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "<u>common element modules</u>" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

a. *The organization and adequacy of the common elements modules is generally well thought out. It seems that the order and prioritization of the first four modules is defensible and covers key elements of systematic review and WOE integration in a chemical-specific assessment.*

b. *In the common elements, EPA should consider adding guidelines for the evaluation of human studies as they are critical to many assessments. The intent of such guidelines would be to add clarity and transparency for how EPA uses human studies, particularly epidemiological studies. Epidemiological studies are mentioned as a part of Module 2, but so critical are these study types to assessments, they deserve stand-alone guidance.*

c. *The guideline modules list NAMs as a part of Module 2. However, given that these are evolving methods that do not have guidelines as yet it may be important to enable them to be represented in a major sub-section of Module 2 or have a separate standalone document to accompany these guidelines, a likely outcome.*

d. *Alternatively, NAMs or high throughput screening and genomic POD's may be better suited to be placed in an endpoint-specific module.*

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

a. *The subsections are well thought out and encompass many traditional endpoints.*

b. *With the adoption of new methods, these subsections may need to be expanded to include the following: epigenetics, HTP screening, genomic PODs or aberrant gene expression changes that serve as POD.*

c. *DART studies should cover endocrine disruption, or it should be a standalone module.*

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing.

a. *The first four modules serve as the core for the guidelines and should be considered first.*

b. *With the paradigm shift to using NAMs, guidelines should cover these areas early in the process so that they may be modified as additional input is given from public and scientific communities over time.*

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

a. *I see further consideration of the use of low-dose extrapolation approaches and methods that harmonize the evaluation of dose-response for cancer and noncancer effects as the top highest priorities of those listed above.*

b. *Various dose-response modeling approaches is important, but of higher importance is how we deal with low-dose or biologically relevant doses.*

The additional consideration of endogenous production of environmental contaminants is important, but represents a rarer event and, as such, should be the lowest priority of the above list.

Dr. Karen Chou

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

Response: The modular approach is carefully considered and constructed. It is a good tool for internal and external communications in setting priorities and task management, especially when dealing with many intertwined and complex issues.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Response:
Module 1: Add premises applied in the principles of risk assessment.
Module 2: This part is well done, nothing to add.
Module 3: In addition to lifestage susceptibility, include other types, such as occupational, sex, and other genetic susceptibility.
Module 4: Some of the new alternative methods may minimize uncertainties, while others may introduce new types of uncertainties into the final assessment. Guidelines for the application of extrapolation factors, including in silico to in vivo, in vitro to in vivo, and organ/tissue-specific to whole animal/human, should be provided to the risk assessors and the public. Existing Uncertainty Factors may need to be redefined and new categories of Uncertainty Factors may be added when the endpoints, such as apical vs. non-apical effects, and dose-response relationship are assessed using new alternative approaches.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

Response: There is no recommendation for changes.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Response: Guidelines for harmonizing cancer and noncancer assessment approaches should be the priority because existing knowledge supports a unified dose-response relationship for cancer-causing and noncancer causing substances. In addition, harmonizing cancer and noncancer assessment approaches could significantly decrease the burden of toxicity testing, reporting, and document review, as well as the number of animals used for testing studies.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

Response:

1. *Low-dose extrapolation is at the center of harmonizing the toxicity assessment for cancer and noncancer health outcomes, therefore, it should be the top priority. This effort should also be made prior to dealing with the model averaging issue.*
2. *Biotransformation products from exposure to exogenous substances have always been a major concern in toxicology and risk assessment. When toxicity will be based on in-vitro and in silico information, EPA should require endogenous products of environmental contaminates be qualitatively and quantitatively identified in target species, including humans, and assessed for potential health hazard.*
3. *Model averaging may be applied to minimize model uncertainty. If the current guidelines do not prevent the application of the model averaging approach, there is no need to set it as a priority item. Supporting resources for the methodology development, however, should continue.*

Dr. Harvey Clewell

Charge to the SAB on the *Consolidated Human Toxicity Assessment Guideline*

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

I believe that the proposed modular approach is a significant improvement over the previous approach, which lacked coherence.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Overall, I believe the list of common element modules is appropriate and includes the key elements of toxicity assessment.

Identifying Planning and Scoping (Module 1) as the initial step in Toxicity Assessment is an important step forward compared to much of the previous guidance on human health assessment and continues the progress made in this area by the IRIS program. Part of this step should also be a preliminary evaluation of Mode of Action based on data for similar compounds identified by QSAR and Read-Across analysis. This can now be easily conducted using apps on the EPA and OECD websites.

The description of Module 2 (Identifying and Evaluating Toxicity Studies) is consistent with the recent efforts by the IRIS program to implement systematic review of toxicity studies.

I am somewhat concerned, however, that the description of Module 3 (Hazard Identification) does not provide a clear statement of the criticality of mode of action (MoA) analysis in the toxicity assessment process. Evidence integration has not typically been performed well in EPA risk assessments. In particular, there has been a tendency to focus hazard identification on the selection of the critical studies that should go forward for dose-response assessment based primarily on a comparison of the associated points of departure, and only apply MoA considerations in the. Despite the emphasis of the current cancer guidelines on the use of MoA evaluation to direct the risk assessment approach, recent assessments have generally failed to adequately incorporate MoA information. There also appears to be an unwillingness to try to apply some form of systematic review to evaluate mechanistic studies, rather than cherry-picking

studies to support going forward with a default approach in the Dose-Response Assessment. This reluctance is certainly driven in part by the potential difficulty of the process, which would involve the review of a wide variety of data, only part of which would be studies conducted on the chemical being assessed. However, the recent inability of the agency to gain NAS acceptance of its toxicity assessments is to a large extent due to the failure to adequately implement a MoA-directed risk assessment approach. In the case of the dioxin cancer assessment, the agency repeatedly resisted NAS requests to show the results of dose-response assessments based on both the linear default and a more scientifically plausible nonlinear approach. This resistance was supported by an evaluation of mechanistic data that appeared to be specifically selected to support the default linear approach, and ignored data to the contrary. Recent risk assessments for arsenic and formaldehyde have also failed to adequately use available data informing the mode of action, and have relied solely on default dose-response approaches, despite strong MoA information supporting alternative approaches. The description of this module needs to provide a clear call for MoA-directed toxicity assessment, regardless of the difficulty of conducting a systematic review of mechanistic data.

In Module 4 (Dose-Response Assessment), it is not clear where inhalation dosimetry (e.g., the 1994 RfC Dosimetry Guidance) fits. Dosimetry is particularly important in the case of aerosol/particle exposures.

There does not appear to be a Module for Risk Characterization. Is that no longer considered to be part of the Toxicity Assessment? I realize that the EPA's position in recent years is that their Toxicity Assessments are not Risk Assessments because they do not include the Exposure Assessment (which they generally have). However, Module 1 would have to include an evaluation of potential exposures in the population of concern in order to put together an appropriate description of the scope and focus of the Toxicity Assessment. Moreover, a Characterization Module is critical to convey the uncertainty in the Toxicity Assessment to the Risk Assessors. Where else could the Characterization go? Are the Risk Characterization Guidelines being withdrawn?

It is crucial that toxicity assessments should include a transparent and comprehensive Characterization Module that is consistent with the OMB Memorandum "Updated Principles for Risk Analysis" (OMB 2007, M-07-24). Important characteristics include:

- Characterizations of risks and of changes in the nature or magnitude of risks should be both qualitative and quantitative, consistent with available data. The characterizations should be broad enough to inform the range of policies to reduce risks.
- Judgments used in developing a risk assessment, such as assumptions, defaults, and uncertainties, should be stated explicitly. The rationale for these judgments and their influence on the risk assessment should be articulated.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific

modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

Modules 5 – 11 represent a good start. Where does general noncancer organ toxicity (liver, kidney, skin, etc.) fit – Other Endpoints?

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Module 3 (Hazard Identification) should be worked on first, to clearly set out the principals of MoA-directed toxicity assessment, including the need for a transparent and objective review of mechanistic data to support an MoA evaluation process that includes relevant studies performed on other chemicals with structural or toxicological similarity.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

a. Use of various dose-response modeling approaches (e.g., model averaging);

Lowest priority.

Bayesian meta-regression (e.g., model averaging) is a powerful approach for analyzing multiple studies, but it is highly susceptible to unintended bias associated with the selection of dose-response models and the definition of quasi-informative prior distributions for model parameters. In addition, due to the unavoidable impact of exposure error in the studies, the observed dose-response can differ significantly from the true dose response, with a tendency toward linearization of the apparent dose-response (Crump 2006, Rhomberg et al. 2011).

An additional challenge with Bayesian meta-regression with epidemiological data is the minimal influence of limited, and often negative, data at low concentrations on the predicted dose-response, which is dominated by the stronger dose-response data at higher concentrations. As a result, even an analysis "within the range of the data" can in fact represent a significant extrapolation below the range of the informative data.

b. Further consideration of the use of low-dose extrapolation approaches;

Highest priority.

c. Additional consideration of endogenous production of environmental contaminants;

Third highest priority.

EPA Guidance currently does not adequately deal with situations where a compound is present endogenously, either as an essential nutrient (e.g., manganese) or as a product of normal metabolism (e.g., formaldehyde, acetone).

d. Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

Second highest priority.

Dr. David Hoel

1) While testifying a few years ago in the Senate to Barbara Boxer and her environmental sub-committee about EPA's handling of TCA, I suggested that EPA use outside experts during their analysis and report development instead of waiting for comments and criticisms from the SAB. This should be done with sensitive materials such as dioxin, formaldehyde etc.

2) EPA uses the linear-no-threshold (LNT) dose-response model in estimating low-dose cancer risk. There has been some suggestion that this may over estimate cancer risks at very low doses especially for radiation. However, recent data analyses and committee reports dismiss this idea. What EPA has not addressed is that bystander effects at low radiation exposures may actually result in the use of the LNT risk model to underestimate the low-dose cancer risks. (see e.g. Brenner et al. Rad Res 155:402-8: 2001). An example is the linear extrapolation of radon lung cancer effects extrapolated from the uranium miner studies. The data is very linear but actually underestimates the low-dose effects observed in residential epidemiology studies by a factor of 4 (see Brenner and Sacks: Int. J. Rad. Biol. 78:593-604, 2002).

3) EPA's IRIS reports include mechanisms, toxicology and epidemiology in developing their cancer risk estimates. Being on a review committee for them on dioxin, I asked if the mechanism/biology section in the IRIS reports had ever impacted the quantitative risk estimates. I was told no except possibly for formaldehyde. Hopefully better use of the non-epidemiology data will impact the quantitative cancer risk estimates that are typically based on epidemiology.

4) For non-cancer effects, safety factors are usually employed with limited epidemiology data. I would like to see a good justification of the particular factors that are used e.g. 10, 20 etc. An example was EPA using a small worker study involving asbestos and plural plaques. Because of being a small unrepeated study the estimated acceptable exposure using a series of large safety factors resulted in a lower acceptable exposure level than that calculated for asbestos and lung cancer. There was of course the argument that plural plaques are a marker of exposure and not an actual adverse health effect.

Dr. Michael Jayjock

My primary expertise is in the evaluation of human exposure in the context of human health risk assessment. As such, I understand that the proper evaluation of hazard or toxic effect is fully half of the risk assessment process. To that end, I have endeavored to study the science of toxicology as it relates to human health risk assessment. That process has caused me to voice opinions and advice to my toxicology colleagues over the years. The strongest effort in that regard is a paper I did with colleagues almost 20 years ago and attached to this email (Jayjock, Lewis and Lynch, Quantitative Level of Protection Offered to Worked by ACGIH Threshold Limit Values Occupational Exposure Limits, AIHA Journal, (62), January/February 2001). This argues for the combination of cancer and non-cancer risk and the use of models to provide quantitative estimates (with uncertainty) of the risk extant at any level of exposure including any exposure limit. Although, not mentioned in this paper, I did suggest, in a subsequent paper, a few year later (Jayjock, How much is enough to accept hormesis as the default?..., Human & Experimental Toxicology, 24, 245-247, 2005) that the emerging science of 'omics would hold the key to actually understanding what might be happening in human tissue at environmentally relevant exposures. It is indeed heartening to see that approach being used within these 2020 draft guidelines.

I was very impressed with the comments and points made by Dr. Fenner-Crisp in response to the charge questions. She has been on the front lines as a very credible, dedicated and capable scientist and public servant relative to these critical issues. I heartily endorse all of her comments, especially her prominent assertion that the NAS become wholly involved at every stage of these deliberations and decisions. From my perspective, the Agency definitely sits within the shadow of public mistrust. I cannot state it better than Dr. Fenner-Crisp:

… Given the lingering concerns about the politicization of the SAB and its committees, it is incumbent upon the agency to engage a broader swath of the scientific community to assure that its outputs reflect an objective view of the state of the science. Consultation with the NAS should begin soon with a conversation similar to that which is occurring now with the SAB and continue at key points along the pathway as illustrated in Figure 1.

Dr. Wayne Landis

I applaud the consideration of probabilistic risk assessment in decision making. By definition, it is not risk assessment unless it is probabilistic. However, there are a number places where the legacy of non-probabilistic approaches exist. One is the continued use of NOAELs and similar measures based on the outcomes of hypothesis testing. The issues with such point estimates can easily be found with a google search. The same can be said of taking a point estimate, even the lower confidence interval, from a regression model. Often, we are attempting to extrapolate to the effects at very low doses because the standard experimental designs are not asking the questions appropriate for risk assessment. Experimental designs need to be altered to answer the key questions in risk assessment, not risk assessment being compromised attempting to accommodate outdated methods.

I have long been a proponent of the use of exposure-response curve fitting instead of hypothesis testing to describe toxicity. I have co-authored several papers on the topic. For a decade I have also worked to integrate causality into a risk analysis and have become increasing skeptical of many approaches claiming to be weight of evidence. I downloaded the EPA guidance document for exposure-response to evaluate its application as a Module.

During my long stint on several SAB subcommittees from the mid 2000s to the early 2010s and with several administrators. I was used to extensive documentation and having time to conduct our own analysis when necessary to answer the charge questions. In comparison this process resembles a rapid screening review than a careful consideration and analysis. I also discourage EPA from referencing documents published behind paywalls in journals in their documentation. Such an example is " (Cumulative risk assessment lessons learned: A review of case studies and issues)". My understanding is that work produced by a U. S. agency cannot be copyrighted by a third party. I also examined several other documents that were available to document different modules.

Consolidated Human Toxicity Assessment Guideline

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

- I consider a modular approach to eventually be a useful approach. It is not clear what the key risk assessments are that these modules support. It appears as if the italicized items in Table 1, Module 2, below are only being considered as factors in a literature review. My experience in evaluating studies for a variety of agencies is that many toxicity studies were conducted as screening studies were a NOAEL or LD50 were the goals. These legacy designs have a number of failings, among them the lack of reported test doses, effects data, minimum effect size, and so on. If the original observations are available,

they may be analyzed using current techniques, but often exposures were not conducted at low doses.

- The approaches are often frequentist in design and analysis. Bayesian statistics, curve fitting and Bayesian networks are being adopted in many other fields and have proven useful. A suggestion to use Bayesian curve fitting was in the 2000 "*Benchmark Dose Technical Guidance Document.*" I have noted that Bayesian curve fitting is now being applied to toxicity data (https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCEA&dirEntryId=343986) and welcome the move.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Module 1

- Modules 1 and 2 do not seem to address ideas of causality-A key first step in any analysis is the construction of at least a proposed cause-effect pathway. Those endpoints based on genotoxicity are likely to have pathways very different from narcosis, endocrine disruption, or interaction with a specific key protein such as AChE. This is one area where the use of adverse outcome pathways (AOPs) may be very beneficial. The relative lack of quantitative AOPs is an issue in making specific predictions, but the framework is a good place to start. When the discussion comes to specific effects it should also bring to mind causal pathways similarities and differences in pathways.
- Be specific on genders and the varied distributions of exposure-response relationships that occur. I do not see that specifically noted in the modules or I missed the implied inclusion.
- Bayesian networks (BNs) and other tools can be applied to evaluate different lines of influence. While BNs are relatively new, the approach can be traced to S. Wright in the early 20th century. Thinking at the beginning of the study of the data analysis framework is key in deciding about the kinds of data that can be addressed.
- The mindset does not address probabilistic factors in so much of the discussion. Toxics and exposure change the probability of physiological responses. Risk –still seems stuck in HQs and divide for a threshold. Even if the HQ is defined probabilistically that limits the kinds of analysis that can be done on the factors contributing to the answer. The tools are now taught to seniors and first year graduate students so they are ready for wider adoption. These tools should also be applied to experimental design and descriptions of causality.

Module 2-I have already identified my preference for exposure-response analysis.

- Module 2 –how do the data from different studies correspond to a proper analysis of exposure response? Again, a conceptual model that describes causality would be useful

here. Given an appropriate analysis and datasets a proper uncertainty and sensitivity analysis can provide insights into the key variables (nodes) in making a prediction.

- Module 2 also seems like a good place for an explicit consideration of an AOP. It is likely that only few key events are necessary to make reasonable predictions of toxicity.
- Dose-response descriptions need to be improved. Even with the BMD approach a point is being presented to describe an exposure response by using the lower confidence interval of the regression. In examining the Guidance Documents is appears that the confidence interval is for the most likely outcome from the dataset. However, the most likely outcome does not necessarily provide protection to the tails of the cause-effect interactions. Prediction intervals should be considered as an additional tool in the decision-making process. Prediction intervals estimate the value of a new observation given the existing model. New observations (effects) can occur far from the boundary of the confidence intervals.
- Discussion of uncertainty and sensitivity in the analysis. In my risk assessment world these are two key characteristics of any evaluation. I do not see a discussion in any of the modules of how important these aspects are in contributing to the risk assessment process.
- Model averaging should take into account prior knowledge—Bayesian model averaging. I am wary of a simple averaging taking place when the outputs of curve-fitting models are discussed. My assumption in this discussion is that Bayesian model averaging is what is being discussed. In this instance weights are assigned to the various model outputs depending on how well they describe the exposure-response relationship. Since the commonly used regression models have little connection to the toxicokinetics of the interaction the equations are more convenience than being based on first principles. It has been demonstrated that when the concentration-response experiment covers the entire range of exposure-response that the various models converge (Moore and Caux 1997). My observation is that experiments designed for hypothesis testing often do not include sufficient observations at doses at which decisions will be made. Hence the different models are not constrained within this region and divergence occurs.
- Testing the accuracy and precision. So, when are we going to test our process for its eventual accuracy and precision? After the modules are produced and in use, how do we know they work?
- The discussion around cumulative effects has been around a very long time. See NRC 2010, Chapter Cumulative effects can be addressed, see NRC 2010. Science and Decisions : Advancing Risk Assessment , Chapter 13, page 213. Note that many of the issues I have discussed in this review are discussed in this keystone publication.

Moore DRJ, Caux P-Y. 1997. Estimating low toxic effects. Environmental Toxicology and Chemistry 16:4 794-801.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

Genders???

Differences in socio-economic class?
Ethnic/Racial differences in access to healthcare, nutrition, etc? will impact susceptibility to stressors.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

It does not seem that there is a clear indication that the experimental design should be amenable to current data analysis tools and that they should describe causality. There has been a growing literature on describing causality and on the fact that the world is both deterministic and probabilistic. The field has advance considerably since the formulation of many of the EPA guidance documents.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

I covered several of these topics above. It is imperative that EPA does a better job of describing exposure-response. I have a few additional comments.

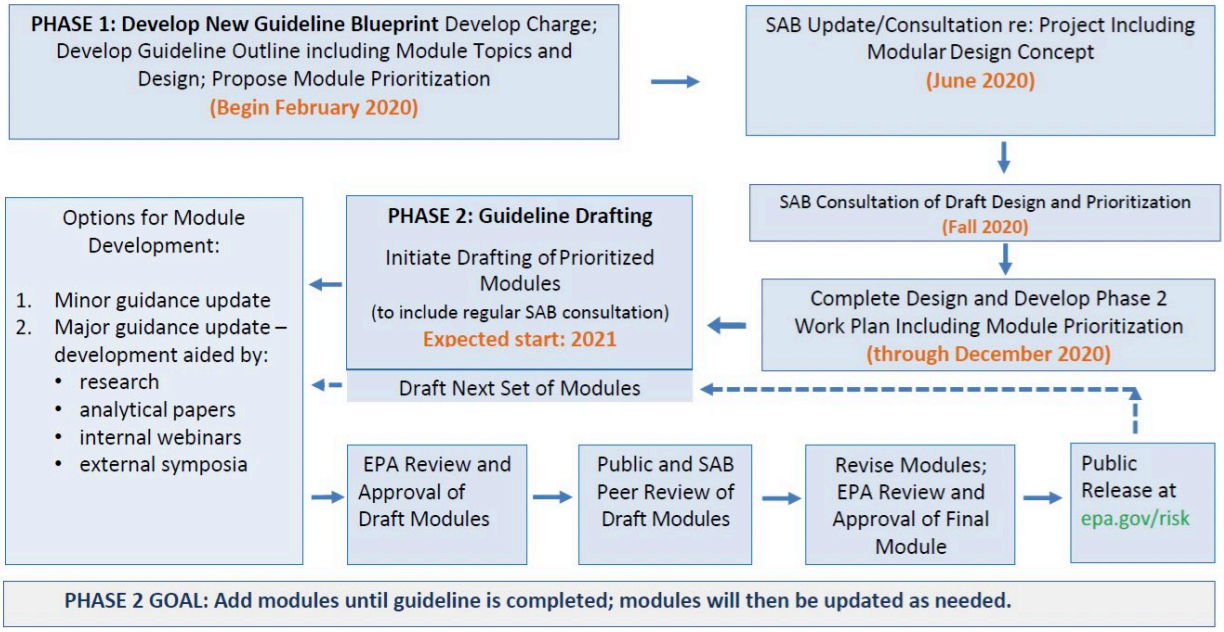
- Use of various dose-response modeling approaches (e.g., model averaging);--*Model average is a tool for reconciling multiple regression assumptions. If the data were adequate throughout the entire exposure-response relationship (in other words, good experimental design) the models should converge. Replication is not as important as having more observations along the exposure-response continuum.*
- Further consideration of the use of low-dose extrapolation approaches-*in other words attempting to have statistical tools save poor experimental design? I am always wary of such attempts; it generates further poorly done experiments. Extrapolation beyond datasets is generally something we teach students as something not to attempt.*
- Additional consideration of endogenous production of environmental contaminants;-(*no comments on this item*) and
- Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects. *First step build conceptual models that describe the various steps in the generation of cancer and non-cancer effects. What are the commonalities? Those commonalities should be the initial steps for consideration of harmonization.*

Table 1: Proposed Modules
Modules are in order of how the Consolidated Guideline could potentially be organized, but not necessarily the order in which they would be written.

Common Element Modules	**Module 1. Planning and Scoping a Human Toxicity Assessment**
These proposed modules would address common elements of an	This module will provide an overview of human health toxicity assessment including key concepts such as fit for purpose, problem formulation, consideration of potential routes of

assessment that pertain to all health endpoints	exposure and overarching considerations including lifestage susceptibility, vulnerable populations and cumulative risk.
	Module 2. Identifying and Evaluating Toxicity Studies This module will cover general principles associated with collecting potentially relevant studies including conducting *a literature search (systematic review), critically appraising different types of data (animal, epidemiological, chamber, modeling, in silico, NAMs, etc.) with respect to study design, power and reliability, data quality evaluation, and identifying data gaps.*
	Module 3. Hazard Identification This module will cover integrating/weighing evidence/synthesizing results across studies, evaluating possible mechanisms/modes of action/adverse outcome pathways including human relevance, and *consideration of lifestage susceptibility.(wgl-of humans or test species??)*

Figure 1: Process/Timeline for Developing EPA's Consolidated Human Toxicity Assessment Guideline



Comments on Figure 1.

Phase 1.

I have a number of questions regarding the diagram.

- What are the goals?
- What kind of accuracy will be required?
- How new are the tools, is this a 21st century process?

- Will the data analysis and decision science be current?
- Are we limited to frequentist approaches to data analysis? I sure hope not. See the references to the Carriger et al. papers below.
- Where is the preliminary conceptual cause-effect framework? Answers to these questions would assist my ability in evaluating the overall process.

Phase 2. How is priority understood if the biggest drivers are not determined in a quantitative fashion? I try to discourage hand-waving.

What are the goals and what are the financial and other design constraints? These societal constraints will limit what the toxicologists and data analysts can do. This can be estimated and provides a context on what EPA is asking the SAB to accomplish.

References

Carriger et al. 2016. Bayesian networks improve causal environmental assessments for evidence-based policy. Environ. Sci. Technol. 2016, 50, 13195−13205. DOI: 10.1021/acs.est.6b03220

Carriger JF, Barron MG. 2016. A practical probabilistic graphical modeling tool for weighing ecological risk-based evidence. Soil and Sediment Contamination: An International Journal, 25:4, 476-487, DOI: 10.1080/15320383.2016.1171293

Dr. Dennis Paustenbach

Consolidated Human Toxicity Assessment Guideline

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

Comments: The modular approach is fine. I would likely agree with others that additional modules under "endpoints" are needed. The key to the success is relatively heavy input by the SAB since this particular panel seems to have all the skills to be able to help EPA achieve their goals. With respect to endpoints, EPA should add an endocrine disruption endpoint as no one seems to have a grip on how to handle it.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Comment: At this time, the general description of module content seems fine. The key is the techniques used to implement.

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

Comment: As noted, EPA should have an endocrine disruption endpoint or simply say "this endpoint is not going to be considered." There is a of confusion about how to handle this one and, personally, I am not sure it is needed at this time. It is extremely difficult to translate endocrine disruption study results (there are a ton of them) vs. dose vs. lack of susceptibility of humans or inability to characterize the hazard vs. naturally occurring substances in the diet (or pharmaceutical agents)

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Comment: I hesitate to propose cancer but it seems to drive the majority of the "priority assessments," let's start there. It will be the most difficult for scientists within EPA and outside EPA to agree upon. I suggest that a subcommittee of SAB meet with the key persons involved in drafting the assessment at various stages in its development. Probably a number of one-day meetings would be a good idea. This approach would make for a lot less work for EPA and SAB when the final proposed document is produced.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

Comment: I agree that this will be a challenge. I like the idea of "model averaging" but that is not so easy. At the 1 in a million risk level, the different models (of almost equal credibility) will yield differences of 1,000 fold. That make true "averaging" an unfair approach. There are ways that those results can be handled more fairly than others.

The background dose or endogenous production of the chemical to be regulated needs to be carefully handled. At the same time, at the doses to be regulated, hormesis needs to be considered. This is important because the regulated community needs guidance as to the kinds of research that they NEED to conduct in order to influence "science" and rulemaking (setting criteria). Perhaps above ALL else, you want the new risk assessment guidelines to inform the regulated community (and universities) as to what studies will be considered (really considered) by the Agency.

It is not an exaggeration to say that very little innovative research has come out of industry or the industry funded research centers in more that 20 (and perhaps 25) years. The bar that EPA set for several decades regarding the weight of evidence needed to convince the agency that a chemical could be safely used, if not consistent with the EPA views, was so great that eventually industry shut down the labs and decided to no longer participate. As one famous CEO said in 1995 or 2000, "We used to spend $500M a year on tox research and studies. We thought the data were convincing and most academics thought the data were more than adequate. We never seemed to convince the Agency and, worse, the Agency's lack of willingness to consider "incremental" regulatory efforts has given us no choice. We will now use litigators rather than scientists for our voice to be heard".

It would be great for science, the agency and the public if this "divide" could be repaired. Clear messages about the studies that are needed would go a long way to help mend the wounds of 25 years of conflict. With the movement away from animal studies, and that needs much more clarification, it would be excellent to see level headed persons write the future guidance documents.

With the pressure to move away from animal testing in less than 10 years, it is important for the Agency to have guidelines as to how "old" animal data will be considered going forward.

Lots to consider.

Dr. Ted Simon

Charge Questions on Toxicity Assessment:

1) This process of guidance development is appropriate to the task. I would, however, include exposure assessment as a tool for prioritization. Dr. John Wambaugh, an EPA staffer in RTP, has written eloquently on this topic and I cite his relevant papers in regard to the NAMs. If a specific chemical can be give a lower priority, smart allocation of the resources for development of toxicity reference values can occur. Perhaps this is included in the "overarching considerations" in Module 1. Maybe a separate module for "exposure prioritization" is needed.

2) As part of module, please include some language that indicate that problem formulation should be viewed as a "voyage of discovery." I found this phrase in the NATO Code of Best Practice for Command and Control Assessment at (https://apps.dtic.mil/dtic/tr/fulltext/u2/a457898.pdf). The point is to ensure the problem formulators keep an open mind.

3) I applaud the idea of an immunotoxicity module but have mixed feelings about including it, as doing so may significantly increase the uncertainty in the process. I would expect most environmental stimuli have some effect on the immune system. The hygiene hypothesis suggests that the current spate of autoimmune disease is due to the elimination of so-called "old friends;" these "old friends" are commensal organisms (invertebrates and protists) from earlier times in human history that provided health benefits and were eliminated as part of a response to other public health goals. The response of medicine now is biologic drugs such as adaimumab or infliximab, monoclonal antibodies against tumor necrosis factor-alpha (TNF-α), a key molecule in the immune response.

   Testament to how little is known about the immune system is the current misperception that pre-existing asthma increases the risk of COVID-19. Whilst both are respiratory diseases, the extant data argues otherwise [1].

   Hence, I would admonish care in developing toxicity factors based on the level of understanding of the portion of the immune system affected. I agree with the goal but a sufficient scientific knowledge base to achieve this goal may not yet be developed. Nonetheless, exploration of immunotoxicity endpoints is worth taking on.

4) I would start module 1 first because lessons learned by doing so may alter the timing and development of the other modules. Thinking hard about planning and scoping should also be a "voyage of discovery."

5) I would agree with all these suggestions. I would prioritize consideration of endogenous production of chemicals, low-dose extrapolation, and harmonization in that order.

6) Regarding the comment made about the so-called key characteristics of carcinogens during the discussion, I strongly disagree with the use of the KCCs. The KCCs were demonstrated to be not better than chance in predicting carcinogenicity. Instead, the use of mode of action as described in EPA's 2005 Guidelines for Carcinogen Risk Assessment and in a huge number of journal articles, e.g., papers by Bette Meek, is a valid means of assessing whether exposure to a particular substance can lead to cancer. KCCs should not be used in the new toxicity assessment guidance.

References

1. Carli G, Cecchi L, Stebbing J, Parronchi P, Farsi A (2020) Is asthma protective against COVID-19. Allergy

Dr. Eric Smith

Discussion/Charge Questions: ***Consolidated Human Toxicity Assessment Guideline***

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

The module approach seems reasonable as a general approach going forward. Having a flowchart might help move through the process.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

Module 1: Should qualitative uncertainties be part of this module?
Module 2: Would it be valuable here to identify critical uncertainties and if there is adequate information to reduce some of these uncertainties. It seems the goal here is to build the framework for the weight of evidence model. Is there a flowchart that would help?
Module 3: Again, it would seem uncertainty plays a role here.
Module 4: This module seems to be much more specific than others. Perhaps this makes it an easier one to complete first. Some of these topics can take a considerable amount of effort (model averaging, probabilistic modeling) to produce useful results in a relatively short document. I presume most of the tools will be frequentist however model averaging and probabilistic modeling can be approached using Bayesian approaches. Will these be considered?

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

The approach based on endpoint-specific modules seems reasonable. Is it worth having study design as part of each of these modules since there is a data interpretation component. Evaluation of strength of evidence is worthwhile (i.e. uncertainties). Is there a need for a module that relates to "strength of conclusions" or how one might combine all the information and evaluate its biological importance? Perhaps a goal would be to identify what is needed to strengthen conclusions. This might include identification of uncertainties, identification of reducible and irreducible uncertainties and how study design might reduce uncertainties.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

This is a difficult question since it depends on some information that I do not have such as budget, personnel available, etc. However, I would suggest working on the modules that are not updates first. Modules 5-11 seem independent of the others so could be done at any time. There is also a need to think about the entire product and a common structure and vocabulary. Providing a common structure for modules that are similar would be a good way to start. The difficulty I see in the project is how much detail should be given about various methods and approaches as there is a large amount of literature about most of the topics.

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- Use of various dose-response modeling approaches (e.g., model averaging);

Model averaging is one method for reducing some of the uncertainty associated with choice of model. There are of course other ways to reduce uncertainty. It would be valuable to give guidance on how much uncertainty can be reduced through model uncertainty. My experience with the method is that it can sometimes provide very good results in some cases however in general there is only a small improvement (say 5-10% in variance reduction). It is not clear how effective it will be for problems such as low dose extrapolation. The document should consider what are the necessary ingredients for successful model averaging and if there are other ways that might also be effective. It is not clear how all of the evidence from various models should be combined to provide an estimate of critical dose levels. This is I think a more important consideration and the document could provide valuable guidance on this topic. Relative to other issues, I view model averaging as low priority. I view data quality and study design as more important issues.

If there is a retrospective study that could illustrate the approaches and compare them to the historical approach this would be a valuable contribution,

- Further consideration of the use of low-dose extrapolation approaches;

This problem is clearly important however it has been the focus of considerable research and discussion. Unless the EPA can provide new information or guidance, I would give it lower priority relative to the next two issues.

- Additional consideration of endogenous production of environmental contaminants; and
- Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.

My only comment here is whether interactions with other contaminants is going to be considered.

Dr. Laura Vandenberg

Discussion/Charge Questions

(1) EPA is planning on using a modular approach to develop its Consolidated Guideline. Please comment on this proposed approach, and if there are other approaches SAB members would recommend EPA consider? This can include comments on Figure 1, Process/Timeline.

There are a number of positive aspects of the proposed modular approach. One is the ability to focus on those areas that require the most immediate attention from EPA, and another is that it allows the individual modules to be updates as new methods, approaches, and priorities are set. The modular approach also allows "new" forms of toxicity to be addressed as they are identified.

(2) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "common element modules" (See Table 1). Comments should include an assessment of each module's description. Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions.

It is difficult to evaluate Modules 1 and 2, because these are actually the most important steps in a risk assessment (based on the NAS's Science and Decisions), and there are insufficient details provided as to whether the modules will be following the best practices as described by the NAS. Does Module 2 intend to cover the steps the NAS refers to as "problem formulation" or is that entirely in Module 1? I think problem formulation and planning & scoping, which the NAS describes as "two[] parallel, stages", should be done as two separate submodules, perhaps in Module 1A and 1B,.

With regard to Module 2, information on whether the systematic review methodologies will follow well-established approaches such as those of the Navigation Guide (developed at University of California SF) or the OHAT method (developed at NTP) is lacking. I do recommend that Module 2 utilize one of these two approaches based on the 2017 NAS review (Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals). It would be very inappropriate for Module 2 to instead rely on the Klimisch score, which is now widely criticized as an insufficient way to evaluate study quality (Ingre-Khans et al. Toxicol Res 2018; Kase et al. Environ Sci Eur 2016).

I also am not convinced by the information provided in Module 2 regarding the evaluation of epidemiological studies. Again, this could be a part of systematic review approaches, but epidemiologic data should also be considered for further analyses (e.g., meta-analyses) and guidance is needed to address these approaches.

I think Module 2 would be improved by breaking it into sub-modules, focused on the evaluation of different "streams" of data. These would include mechanistic data (in silico, in vitro, NAMs); in vivo laboratory animal data; and observational data (epidemiology and wildlife).

Module 3 should address specifically how Adverse Outcome Pathways will be utilized, especially when the data that are available (e.g., in a specific case study) examine molecular initiating events and/or key events, but do not examine the adverse outcome (e.g., because of the use of NAMs). At this point, very few AOPs have been accepted by EAGMST / OECD, and I am not aware of the EPA's use of AOP data in any risk assessments, so additional detailed guidance, as well as some illustrative examples is needed for this module.

What is also unclear from this approach is how EPA will address incongruence between epidemiology studies and in vivo laboratory animal studies. By this, what I mean is that guideline studies have proven to be poorly predictive of the effects, or the doses, at which harm is observed in human populations. Guideline studies examine toxicity outcomes, but rarely do these outcomes tell us what we need to know about disease (as a hazard, or as a risk). There are now countless stories where epidemiology studies (and I mean reproducible effects, observed across cohorts and in different populations) show that "typical" human exposures are associated with disease and/or dysfunction, yet exposure levels are below the RfD/ADI. How can this be? What is clear is that guideline studies, and the application of the test guidelines, is insufficient to protect human health. When the data clearly show this, it is too late --- once epidemiology studies have been conducted, exposures have already occurred and harm has come to at least a subset of the population. How is this addressed by these modules?

(3) Please comment on the scientific adequacy, completeness, organization and other relevant considerations regarding EPA's list of proposed "endpoint-specific modules" (See Table 1). Any recommendations for new, expanded, consolidated or split modules should come with suggested descriptions and other relevant guidance.

It is clear that a module on "endocrine disruption" is needed, and I recommend that this go beyond the guidance that was offered in the development/implementation of the EDSP. In the past twenty years, it has become clear that endocrine toxicity goes beyond EATS, and the failure to consider modern knowledge of the endocrine system will put the public at risk.

There is insufficient detail to evaluate the organ specific toxicity modules.

(4) EPA will need to set priorities and start some modules before others. What modules would SAB members suggest EPA work on first and why? This may include commentary on the extent of update needed for each of the existing guidelines.

Although I do not think it is necessary or appropriate to prioritize the modules in the order that they are presented, I do think the first two modules are in need of the most immediate attention. Again, Science and Decisions suggests that Planning and Scoping, and Problem Formulation,

are the most important steps in a risk assessment; it also notes that these two parallel steps are "critical, but often underused". For this reason, Module 1 should receive the most immediate attention. Frankly, without setting the stage for how/why the remainder of the steps will progress, the assessment is doomed to fail.

I also suggest addressing Module 2 as a top priority. As noted above, methods such as the Navigation Guide or the OHAT method are available, and have already been validated (and evaluated by the NAS). I recommend utilizing one (or both) of these approaches as the starting point for Module 2.

After that, another priority should be the development of the Immunotoxicity guidance (because no guideline currently exists) and the proposed module on endocrine toxicity (because the current approaches are insufficient).

(5) EPA received many comments from SAB members on dose-response issues. Comments that came up multiple times include those shown below. Please comment on which of these or other issues SAB members would consider to be of higher priority:

- **Use of various dose-response modeling approaches (e.g., model averaging);**
- **Further consideration of the use of low-dose extrapolation approaches;**
- **Additional consideration of endogenous production of environmental contaminants; and**
- **Methods that would harmonize the evaluation of dose-response for cancer and noncancer effects.**

I would strongly suggest addressing methods that could harmonize cancer and non-cancer effects. It is clear that the belief in thresholds is not based on data, and that thresholds can rarely be demonstrated empirically (and are often not observed in populations). Thus, harmonization of approaches for cancer and non-cancer effects, avoiding the indefensible use of thresholds, would be an improvement in dose response issues.