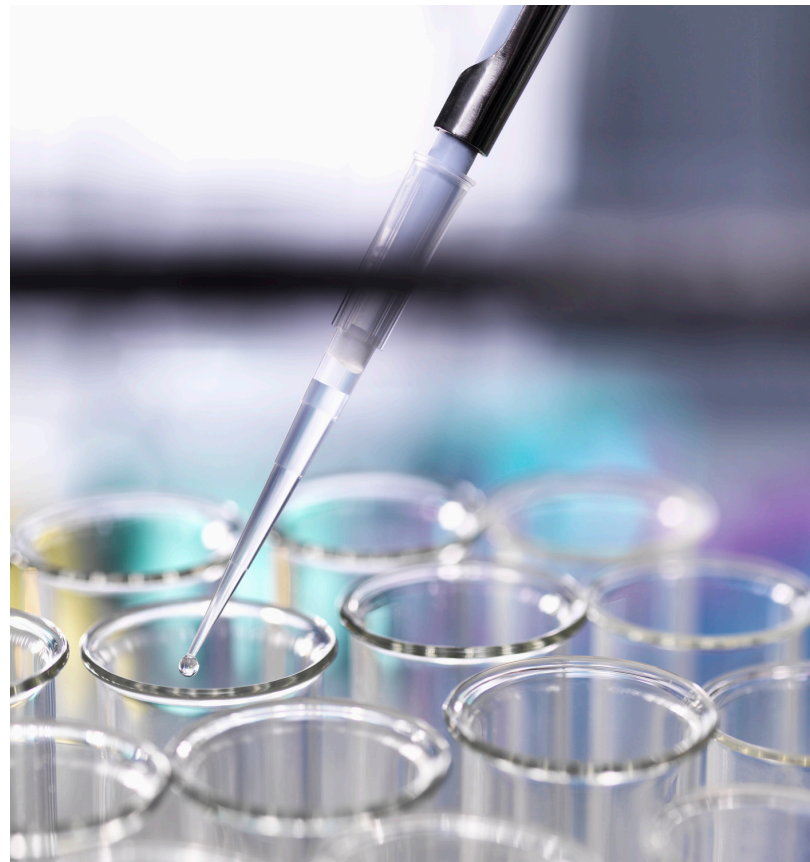# Review of EPA's 2022 Draft Formaldehyde Assessment

Jonathan M. Samet, M.D., M.S.,
Study Chair

August 8, 2023
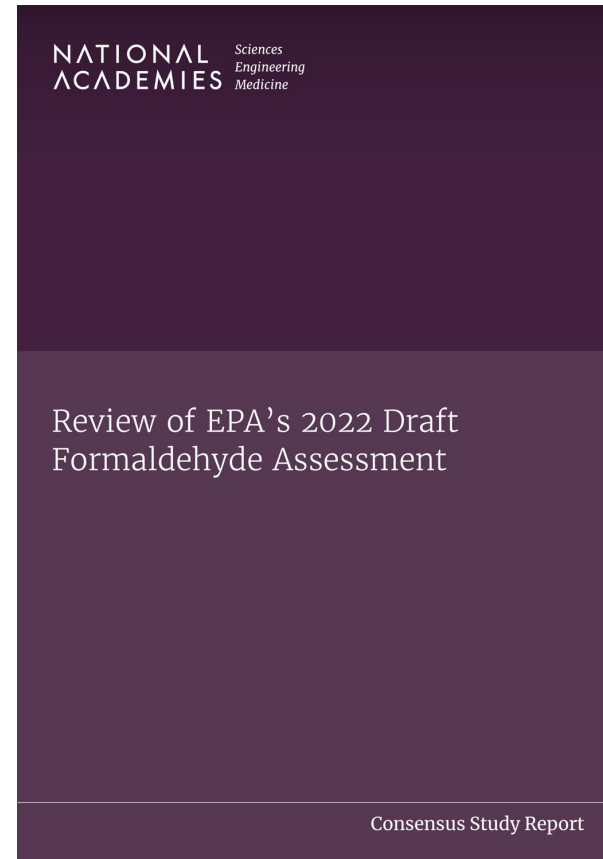
NATIONAL ACADEMIES *Sciences Engineering Medicine*

Our new report

Committee on *Review of EPA's 2022 Draft Formaldehyde Assessment*

Sponsor:

U.S. Environmental Protection Agency

Board on Environmental Studies and Toxicology

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Review of EPA's 2022 Draft Formaldehyde Assessment

Consensus Study Report

Committee Members

Jonathan Samet , MD, MS (*Chair*) (NAM)
Colorado School of Public Health

Aisha S. Dickerson, PhD, MSPH
Johns Hopkins University

Dana C. Dolinoy, PhD
University of Michigan School of Public Health

David C. Dorman, DVM, PhD
North Carolina State University

Rakesh Ghosh, PhD
University of California, San Francisco

Sabine S. Lange, PhD
Texas Commission on Environmental Quality

Andrew F. Olshan, PhD
University of North Carolina

Ivan Rusyn, MD, PhD
Texas A&M University

Lianne Sheppard, PhD
University of Washington School of Public Health

Katya Tsaioun, PhD
Johns Hopkins University

Joseph L. Wiemels, PhD
University of Southen California

Lauren Zeise, PhD
California Environmental Protection Agency

Yiliang Zhu, PhD
The University of New Mexico

NATIONAL ACADEMIES *Sciences Engineering Medicine*

NASEM Staff

Kathryn Z. Guyton, PhD
Senior Program Officer and Project Director

Elizabeth B. Boyle, MPH
Senior Program Officer

Anthony DePinto, MPH
Associate Program Officer

Brenna Albin, BS
Senior Program Assistant

Darlene Gros, BA
Senior Program Assistant

Clifford S. Duke, PhD
Board Director, Board on Environmental Studies and Toxicology

Report Reviewers:

A. John Bailer
Miami University

Rick Corley
Greek Creek Toxicokinetics Consulting, LLC

Susan Felter
Procter & Gamble

Penny Fenner-Crisp
Independent Consultant

Julie Herbstman
Columbia University; Columbia Center for Children's Environmental Health

Juleen Lam
California State University East Bay

Malcolm Macleod
University of Edinburgh

Moshe Oren (NAS, NAM)
Weizmann Institute of Science

Sue Junks-Robertson (NAS)
Duke University

Joe Rodricks
Ramboll

Review was overseen by:
Danny Reible and **Don Mattison**

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Study Timeline

Meeting 1: October 12-13, 2022

Meeting 2: December 22, 2022

Meeting 3: January 30-31, 2023

Meeting 4: March 3, 2023

Meeting 5: April 5, 2023

Meeting 6: April 12-13, 2023

Meeting 7: May 10, 2023

Meeting 8: June 14, 2023

Meeting 9: July 31, 2023

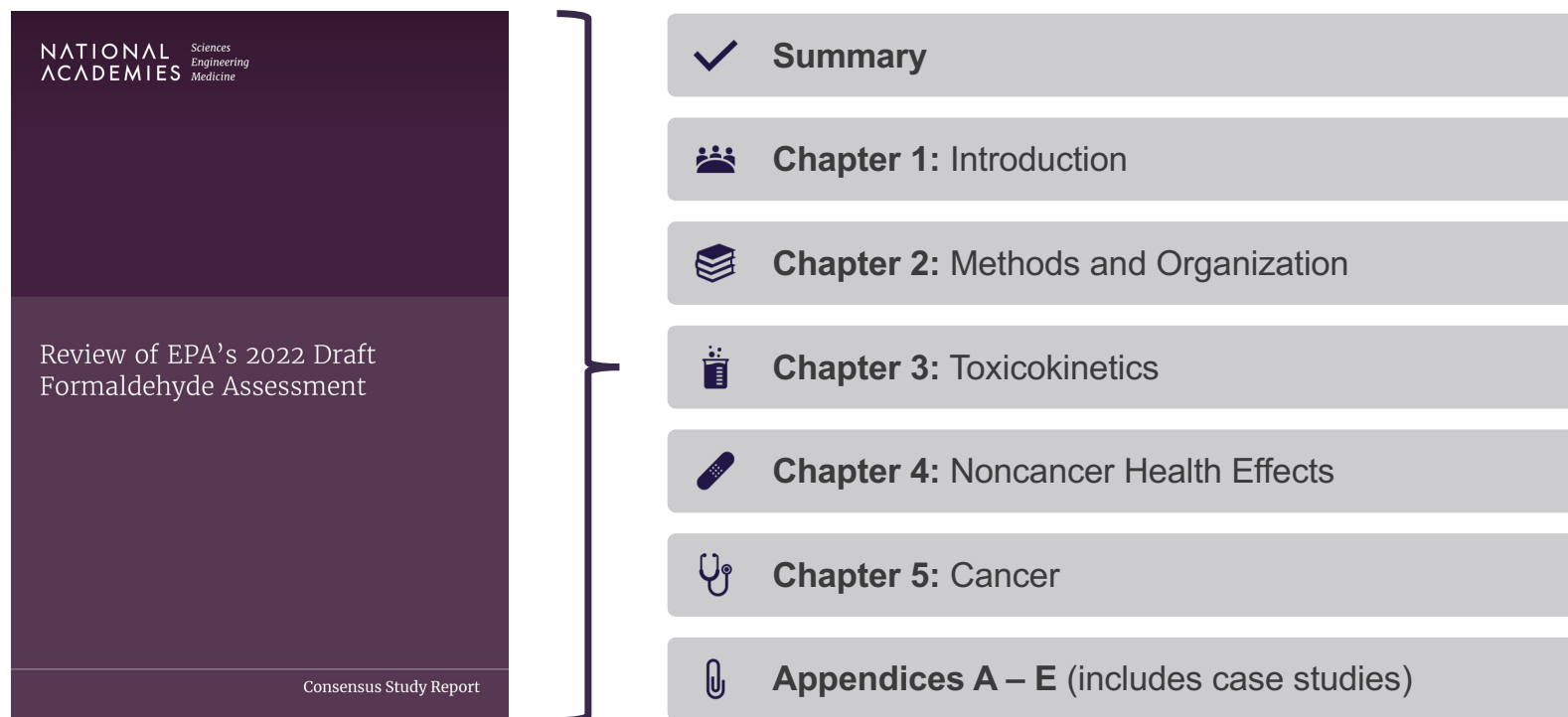**Prepublication report on NAP.edu: August 8, 2023, 11:00 AM**

Committee's Statement of Task

- Conduct a scientific review of EPA's draft document referred to as the Integrated Risk Information System (IRIS) Toxicological Review of Formaldehyde, plus appendices

- Assess whether EPA's draft document adequately and transparently evaluated the scientific literature, used appropriate methods to synthesize the current state-of-the science, and presented conclusions regarding the *hazard identification analysis* and *dose-response analysis* of formaldehyde that are supported by the scientific evidence

- The committee **will not** conduct its own hazard assessment of formaldehyde, **nor** will the committee address the broader aspects of the IRIS Program.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Committee's Statement of Task

- Recommendations about the IRIS assessment will be prioritized as follows:

    – **Tier 1**: recommended revisions that are important for EPA to consider and address to improve critical scientific concepts, issues, or narrative in the assessment.

    – **Tier 2**: suggested revisions that are encouraged to strengthen or clarify the scientific concepts, issues, or narrative in the assessment but are not critical. Other factors, such as agency practices and resources, might need to be considered by EPA before undertaking the revisions.

    – **Tier 3**: considerations that might inform future evaluations of key science issues or inform development of future assessments.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Organization of the Report

Review of EPA's 2022 Draft Formaldehyde Assessment

Consensus Study Report

- ✓ **Summary**
- **Chapter 1:** Introduction
- **Chapter 2:** Methods and Organization
- **Chapter 3:** Toxicokinetics
- **Chapter 4:** Noncancer Health Effects
- **Chapter 5:** Cancer
- **Appendices A – E** (includes case studies)

This presentation only includes *selected* highlights and recommendations from the full report.

# Development of the EPA Formaldehyde Draft Assessment

EPA's development of the formaldehyde Draft Assessment in the context of reports from:
- NRC (in 2011 and 2014)
- NASEM (in 2018 and 2022)
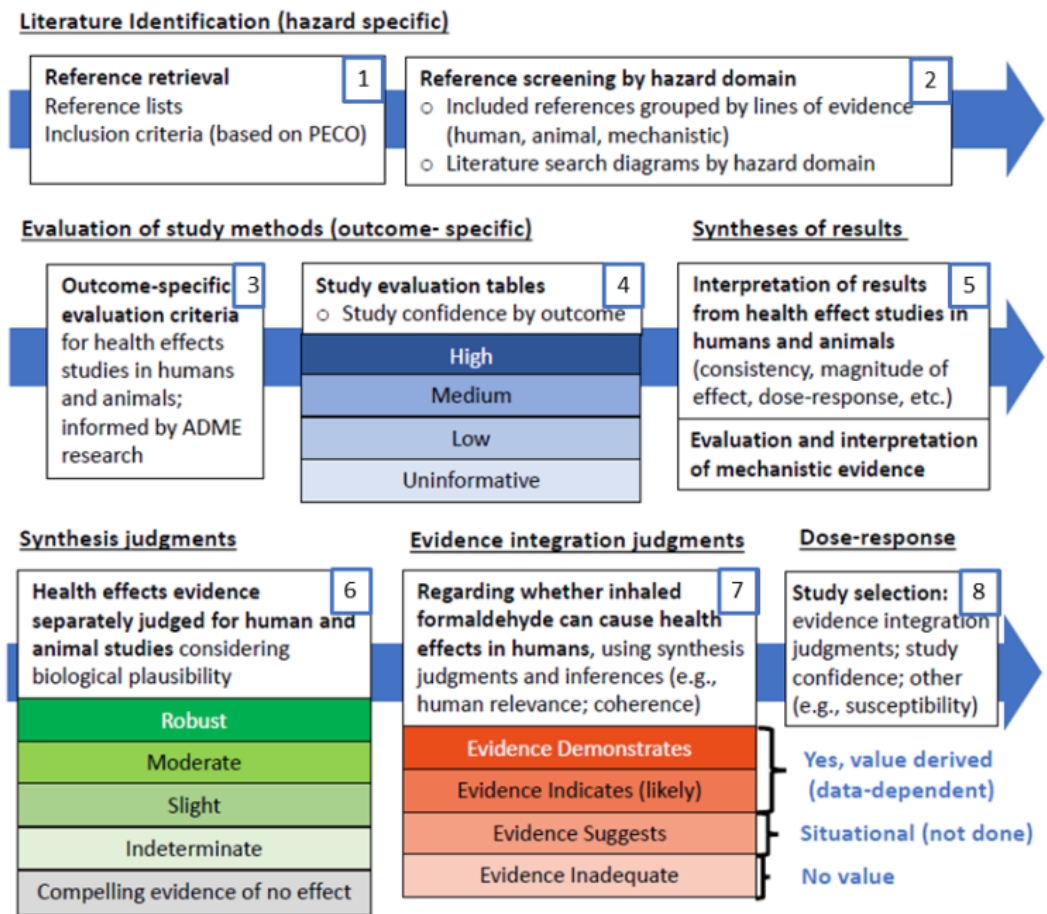
NASEM reports encouraged the IRIS Program to:
- adopt systematic review methods
- develop a staff handbook with guidance on the methods used in IRIS assessments
- develop an a priori protocol for each major IRIS assessment

Figure 1-2. Timeline of development of EPA's formaldehyde Draft Assessment
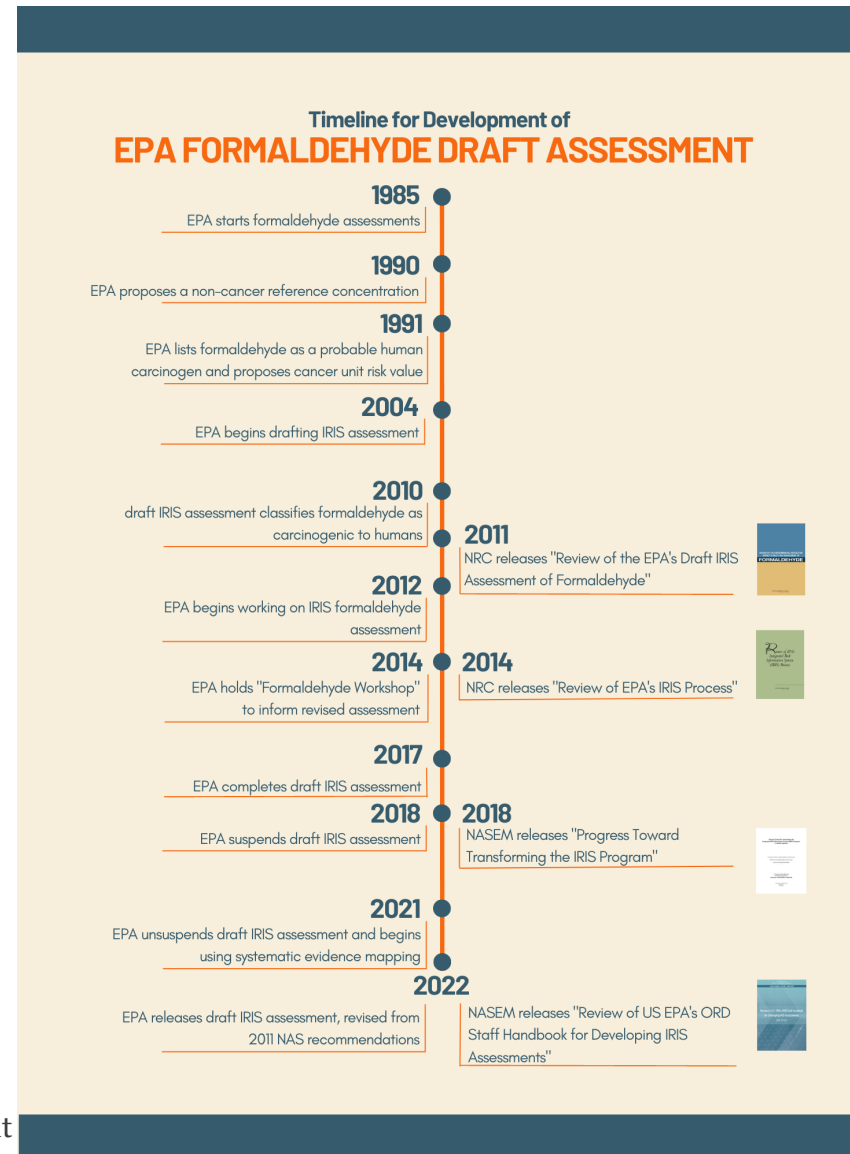
Overview of Committee's Approach

- Evaluated EPA's protocols against accepted systematic review methods
- Critiques and suggestions offered alongside each step in the assessment (Figure 1-3)
- Case study approach for transparency and replicability
- Relied on documentation provided by EPA and in their responses to the committee's queries

Literature Identification (hazard specific)

Reference retrieval [1] Reference lists Inclusion criteria (based on PECO)	**Reference screening by hazard domain** [2] ○ Included references grouped by lines of evidence (human, animal, mechanistic) ○ Literature search diagrams by hazard domain

Evaluation of study methods (outcome- specific)   **Syntheses of results**

Outcome-specific evaluation criteria [3] for health effects studies in humans and animals; informed by ADME research	**Study evaluation tables** [4] ○ Study confidence by outcome High Medium Low Uninformative	**Interpretation of results from health effect studies in humans and animals** [5] (consistency, magnitude of effect, dose-response, etc.) **Evaluation and interpretation of mechanistic evidence**

Synthesis judgments   **Evidence integration judgments**   **Dose-response**

Health effects evidence separately judged for human and animal studies considering biological plausibility [6] Robust Moderate Slight Indeterminate Compelling evidence of no effect	**Regarding whether inhaled formaldehyde can cause health effects in humans**, using synthesis judgments and inferences (e.g., human relevance; coherence) [7] Evidence Demonstrates Evidence Indicates (likely) Evidence Suggests Evidence Inadequate	**Study selection:** [8] evidence integration judgments; study confidence; other (e.g., susceptibility) Yes, value derived (data-dependent) Situational (not done) No value

Figure 1-3. Systematic review approach used by EPA to complete the 2022 Draft Assessment. Note: Modified from EPA's presentation to the committee on October 12, 2022.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Responsiveness to Prior Recommendations and Documentation of Methods

Overall, the committee found that the 2022 Draft Assessment **is responsive to the broad intent** of the 2011 NRC review of EPA's 2010 Draft Assessment and the 2014 NRC review of the IRIS process

Methods spread across all 3 documents of the 2022 Draft Assessment:

- Main Assessment (789 pages)

- Appendices (1059 pages)

- Assessment Overview (192 pages)

Recommendation 2.1 (Tier 1): EPA should revise its assessment to ensure that users can find and follow the methods used in each step of the assessment for each health outcome. EPA should eliminate redundancies by providing a single presentation of the methods used in the hazard identification and dose-response processes. A central roadmap and cross-references are also needed to facilitate access to related sections across the different elements of the assessment (e.g., appendixes, main document) for the different outcomes analyzed. Related Tier 2 recommendations would amplify the impact of this Tier 1 recommendation in improving the assessment.

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Responsiveness to Prior Recommendations and Documentation of Methods

Recommendation 2.2 (Tier 2): In updating the assessment in line with the Tier 1 Recommendation 2.1, EPA should further clarify the evidence review and conclusions for each health outcome by giving attention to the following:

- Using a common outline to structure the sections for each health outcome in order to provide a coherent organization that has a logical flow, by
  - adding an overview paragraph to guide readers at the start of sections for each of the various health domains, and
  - including hyperlinks to facilitate cross-walking among sections within the document;
- Moving lengthy, not directly used information to an appendix;
- Including a succinct executive summary in the Main Assessment; and
- Performing careful review and technical editing of the documents for consistency across the multiple parts of the Draft Assessment, including across the Assessment Overview and Appendices. (The Assessment Overview could be entirely removed if the above recommendations were carried out.)

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Evidence Identification (Steps 1 and 2)

- Generally, the committee found the literature searches to be adequate
  - the approaches used were **consistent with the state of practice at the time**

- The **origins** of the various population, exposure, comparator, and outcome (PECO) statements are **less clear** (noncancer examples below)

Noncancer outcome	Example
Allergy and asthma	age cutoffs are not stated and clearly applied
Sensory irritation	rationale for excluding outdoor air studies is unclear
Respiratory pathology	search terms and inclusion and exclusion criteria used to search human and animal evidence are discrepant

Recommendation 2.3 (Tier 2): EPA should expand the text explaining the choices of the elements of the PECO statements.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

13

Study Evaluation (Steps 3 and 4)

- Overall, outcome-specific criteria for evaluating studies were appropriate

- However, the information is presented in several different locations in the documents, and in some cases is **inconsistently presented and integrated** across them

- The committee's evaluation revealed **inconsistencies between the stated criteria and the rationale** for conclusions on study confidence

  o Evident in EPA's evaluation of human and animal studies across noncancer outcomes

Recommendation 2.4 (Tier 2): EPA should thoroughly review the Draft Assessment documents to address issues of consistency and coherence so as to ensure that its methods can be applied and replicated with fidelity. The reviews for each outcome in Chapters 4 and 5 provide more specific guidance.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Evidence Synthesis (Steps 5 and 6) and Integration (Step 7)

- The strength-of-evidence categories and how they were applied to the overall evidence judgments are **generally clear and appropriate** for the human and animal evidence streams

- The 2022 Draft Assessment **deviates from established methods for inferring causation** in several respects:
  - blurring of the boundary between evidence synthesis and integration
  - the choice of terminology used to describe the strata in the four-level schema for classifying strength of evidence

Recommendation 2.5 (Tier 2):
- The assessment should be edited to more sharply demarcate the synthesis and the integration of evidence discussions.
- EPA should expand the narrative descriptions of the evidence integration step, or should follow published methodology while providing detailed explanation of any adaptations.

Evidence Synthesis (Steps 5 and 6) and Integration (Step 7) – Mechanistic Evidence

- The committee found that EPA is **thorough and transparent** in identifying the relevant information for mechanistic evidence

- Similarly, the term "**other inferences**" is used in the sections on integration of cancer evidence, but is not explained

- However, the definition of "**impactfulness**" and how this concept was applied are not well described

Recommendation 2.6 (Tier 2): To increase the transparency of the evaluation of mechanistic data, EPA should clarify key terms (e.g., "impactfulness," "other inferences") and their application to specific studies. "Impactfulness" can be defined (in Table F-12 and elsewhere), and "other inferences" can be explained in discussing the approach to evidence integration in the "Preface on Assessment Methods and Organization."

Evidence Synthesis (Steps 5 and 6) and Integration (Step 7) – Noncancer and Cancer Judgments

Recommendation 4.7 (Tier 2): EPA should clarify the basis for its synthesis judgments and provide additional information about the studies on which they are based, such as the formaldehyde levels observed, as well as the exposure ranges or other measure of variability. The study summary tables (Tables 1-6 to 1-9) should be updated to provide an organized distillation of the points made in the evidence synthesis text.

Recommendation 5.1 (Tier 2): While the narrative describing the application of criteria for each site is well done, EPA should enhance clarity by providing explicit statements in section 1.2.5 summarizing synthesis judgments for each criterion (consistency, strength, temporal relationship, exposure-response relationship, etc.).

Dose-Response Assessment (Step 8)

- The committee found the considerations for selection of dose-response studies to be **reasonable**

- EPA provides criteria for study inclusion in the dose-response assessment, but do**es not include any discussion of how these criteria were applied** to the specific studies chosen for dose-response
  - Test case: **Hanrahan et al. (1984)**

Recommendation 4.6 (Tier 2): EPA should clarify and clearly state the criteria used to select the studies for dose-response analysis of noncancer endpoints

Dose-Response Assessment (Step 8) – Derivation of the RfC

Recommendation 4.16 (Tier 2): EPA should carefully address the following points regarding the derivation of the RfC:

- Fully disclose data extracted from original study reports using HERO or other means.
- Cite relevant guidance documents regarding the use of a mean versus median and arithmetic mean versus geometric mean to estimate a lowest observed adverse effect level or no observed adverse effect level.
- In reanalyzing data from published studies, the use of raw data is preferred. Aggregated data may be used when appropriate. At a minimum, group size, group mean, and a measure of variance (e.g., group standard deviation or standard error of the mean) for each exposure level are needed to capture data variation in a reanalysis of dose-response.
- Avoid fitting a dose-response model that has as many parameters as the number of distinct aggregated data points taken from the published literature. Report and consider only models that meet the goodness-of-fit criterion EPA accepts.
- To ensure that the resulting benchmark concentration lower bound is not artificially overestimated, better account for within-group variability in the dose-response analysis of Hanrahan et al. (1984) to address limitations arising from reliance on only secondary, aggregated rates per exposure group that were extracted from the plot of the originally fitted model.
- Be more explicit as to how the final RfC was chosen (in Figure 2-2 of the 2022 Draft Assessment and elsewhere).

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Dose-Response Assessment (Step 8) – Cancer Endpoints

- The analyses generally **followed the process** outlined in the 2022 IRIS Handbook and were consistent with the 2005 Guidelines for Carcinogen Risk Assessment

- The decision points and analyses were also **responsive to the recommendations** of the 2011 NRC committee
  - Documented in Appendix D

- Specific recommendations regarding the cancer dose-response assessment concern:
  - criteria for study selection
  - procedure and justification for pooling the data from two animal studies into one analysis
  - discussion of uncertainties and variabilities
  - characterization of inhalation unit risk estimate

Recommendation 5.4 (Tier 2): While the criteria for selecting the Beane Freeman et al. (2013) study can reasonably be discerned from the 2022 Draft Assessment, EPA should provide clearer statements of the criteria and comparison of studies with such criteria, in tabular format, to improve transparency and clarity. EPA should add to such a table other studies that evaluated the same cancer outcome so it is apparent why the selected study was superior for the purposes of dose-response analysis.

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Concluding Remarks: The Path Forward

The revised 2022 Draft Assessment follows the advice of prior National Academies committees, and its findings on hazard and quantitative risk are supported by the evidence identified.

At present, the description of methods in several places throughout three lengthy documents is perhaps the most critical area for structural and editorial revisions.

Implementation of the committee's recommendations would strengthen EPA's conclusions on the many noncancer outcomes reviewed, as well as the cancer hazard identification and dose-response conclusions.