# EPA Responses to NASEM panel questions for the January 30, 2023 public meeting

Date: February 8, 2023

To: Kathryn Z. Guyton
Senior Program Officer, National Academies/BEST

On January 6, 2023, EPA received a request from the NASEM panel tasked with reviewing the draft IRIS formaldehyde (inhalation) assessment to provide written responses to the following four questions (and sub-questions) for discussion at a public meeting on January 30, 2023. In some instances, the interpretation of the question as written was not completely clear to EPA, or seemed to overlap with other questions (e.g., questions 1 and 2), so it is possible that the information the panel was interested in receiving in response to one question might be found in response to a separate question. In addition, based on the follow-up questions from the panel at the January 30th meeting, several minor elaborations and clarifications not initially included within this response document were added, and this updated version of the responses was provided to NASEM on February 8, 2023. To ensure transparency for the public, yellow highlighting is used to indicate adjustments (including grammatical corrections) made to the original text EPA provided to NASEM on January 27, 2023.

Thank you for the opportunity to address these questions. We understand these questions and responses will be made available to the public.

Kristina Thayer, Ph.D.
Director, Chemical & Pollutant Assessment Division (CPAD)
Center for Public Health & Environmental Assessment (CPHEA)
Office Research and Development, U.S. EPA

1. Are there specific systematic review protocols that were used for the assessment methods depicted in the overview of the IRIS approach as presented by EPA (slide 12 of EPA's presentation, below [Figure 1])?

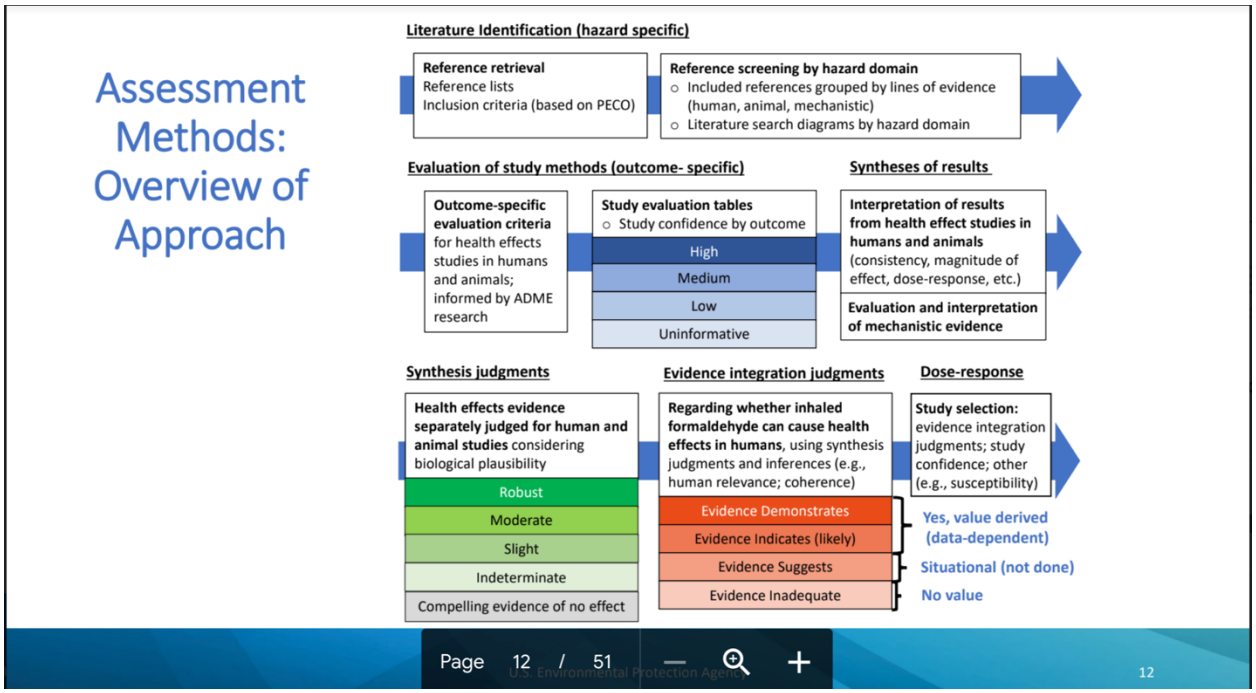


*Figure 1. EPA Figure on Overview of Assessment Methods Referenced in Questions from the NASEM Panel ("Slide 12")*

EPA Response to Question 1 (general):

In the draft IRIS formaldehyde assessment, the "Preface on Assessment Methods and Organization" describes the methods and approaches applied to develop the assessment, with some additional and more granular, health effect-specific methodological considerations provided in the Appendices. The description of the assessment methods provided across the Preface and Appendices of the draft formaldehyde assessment constitutes the level of detail that would be presented in a modern-day IRIS assessment protocol. That stated, because these methods were not separately released in advance of releasing the assessment (the current practice of the IRIS Program for protocols), this content is not referred to as an assessment protocol. We note that inclusion of the assessment methods within the assessment documents rather than in a separate protocol is consistent with the practices within the IRIS Program at the time the formaldehyde assessment was being developed during 2012-2017 (see Addendum and Figure 2). Although a protocol was not separately released

prior to the draft assessment, the description of methods included within the draft assessment has been available for comment throughout the 7-Step IRIS Process.

For additional context, the draft formaldehyde assessment had already completed Step 1 of the IRIS process (draft development) when the IRIS Program first established the practice of releasing protocols for newly initiated assessments, with the first IRIS protocol released in 2018 (see Figure 2). However, at that time, further advancement of the draft formaldehyde assessment through the IRIS process (Step 2 and beyond) was suspended at the request of EPA leadership; the assessment was unsuspended in 2021 (https://www.epa.gov/sites/default/files/2021-03/documents/iris_program_outlook_mar2021.pdf). Given that a complete Step 1 draft had been developed by 2017, EPA did not consider it appropriate (or a pragmatic use of resources) to develop and release a protocol for a finished draft.

The presentation of the methods as part of the draft assessment materials builds from the IRIS "Preamble" approach and is consistent with advice from NAS panels in 2011 and 2014.[1] The "Preamble" approach was first applied within the draft ammonia and TMBs assessments in 2012, and evolving aspects of the systematic review methods were included within the front matter of draft assessments, as demonstrated in subsequent public draft assessments of benzo[a]pyrene (2014), RDX (2016), and ETBE and TBA (2017), as detailed in the Addendum. Draft assessments at this time were growing and evolving in their presentation and handling of evidence synthesis and integration methods, and none included a separately released protocol. In fact, the methods provided in the 2017 draft IRIS formaldehyde assessment that had completed Step 1 of the IRIS process represented the first draft IRIS assessment to describe and incorporate such methods for every component of draft development, including evidence synthesis and integration.

In some instances, updated methods were applied and documented in the draft formaldehyde assessment for components of draft development as those methods evolved (e.g., the use of systematic evidence map methods to update the draft from 2017

---

[1] NAS (2011) (p. 14): "Chapter 1 of the draft assessment needs to discuss more fully the methods of the assessment. The committee is recommending not the addition of long descriptions of EPA guidelines but rather clear concise statements of criteria…" https://nap.nationalacademies.org/catalog/13142/review-of-the-environmental-protection-agencys-draft-iris-assessment-of-formaldehyde

NAS (2014) (p.5) "The preamble is a useful statement, which will presumably be updated as methods and procedures are modified and updated, but it does not substitute for an overview that indicates how the general principles in the preamble have been applied in any given assessment." (p. 6) "EPA should include protocols for all systematic reviews conducted for a specific IRIS assessment as appendixes to the assessment." https://nap.nationalacademies.org/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process

to 2021; consistency in presenting and documenting evidence synthesis and integration judgments). However, most systematic review components represent earlier workflows of the currently used approaches (e.g., literature screening and study evaluations conducted around 2012–2015 were performed without the use of specialized software tools currently used by the IRIS Program to simplify transparent documentation), although the underlying concepts and considerations were the same. This incremental incorporation of methods follows the NAS (2011) recommendation that the IRIS Program should not stop producing assessments to wait until the Program has optimized methods for developing those assessments in a systematic and transparent manner (i.e., the IRIS Handbook).[2]

As detailed in the Addendum and other responses below, although a protocol was not released prior to releasing the draft IRIS formaldehyde assessment, the methods developed and applied in producing the draft (and released with the assessment) are consistent with the IRIS assessment development methods used currently. Although the IRIS Handbook was only released recently (December 2022), the methods described in the Handbook (and applied in developing the IRIS formaldehyde assessment draft) have been publicly available for several years through a variety of outlets:

- Evolving methods in other draft and final IRIS assessments released between 2014-2021.
- Presentations at EPA-sponsored and external workshops on systematic review and formaldehyde specifically (Addendum Table A-2).
- Publications on systematic review methods (e.g., Addendum Table A-3).
- Publicly available materials provided to and cited by the NAS 2014 committee.
- Presentation materials provided to and cited by the NAS 2018 committee[3] (see Appendix C of that report). This presentation outlined detailed methods for each step of systematic review in IRIS assessments and was coauthored and presented by members of the IRIS formaldehyde assessment team.

---

[2] NAS (2011) (p. 151): "Although the committee suggests addressing some of the fundamental aspects of the approach to generating the draft assessment later in this chapter, it is not recommending that the assessment for formaldehyde await the possible development of a revised approach."

[3] https://nap.nationalacademies.org/catalog/25086/progress-toward-transforming-the-integrated-risk-information-system-iris-program (NASEM, 2018 IRIS review)

Although this occurred after the draft IRIS formaldehyde assessment was suspended, it provided detailed assessment methods being applied in all drafts, including the draft IRIS formaldehyde assessment.

- Methods in IRIS protocols released between 2018–2021. Approximately eight protocols were released while the draft IRIS formaldehyde assessment was suspended (https://www.epa.gov/iris/iris-recent-additions), each of which was publicly released for comment (and some for external peer review) and included template language on the systematic review methods for each step of assessment development.

- Most notably, the methods used to develop the draft IRIS formaldehyde assessment were foundational to the development of the methods presented in the IRIS Handbook[4], posted by EPA in December 2022 after being revised in response to a NASEM peer review report. The Handbook, which included team members from the formaldehyde team as lead authors, was publicly released for comment in 2020, prior to the release of the draft IRIS formaldehyde assessment.

Overall, the description of the methods applied in developing the draft IRIS formaldehyde (inhalation) assessment is provided as a Preface to the Toxicological Review and Appendices, consistent with the evolving systematic review methods that were being applied over time within the IRIS Program (see Figure 2). Additional information potentially relevant to this question is provided in response to Question #2 below, and the specifics of the methods used in the current draft IRIS formaldehyde assessment are outlined in responses to [1a] through [1c] below.

[4] https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=545991 (posted Handbook, 2022); https://nap.nationalacademies.org/26289 (NASEM report on Handbook, 2022); https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=541571 (draft Handbook, 2020)

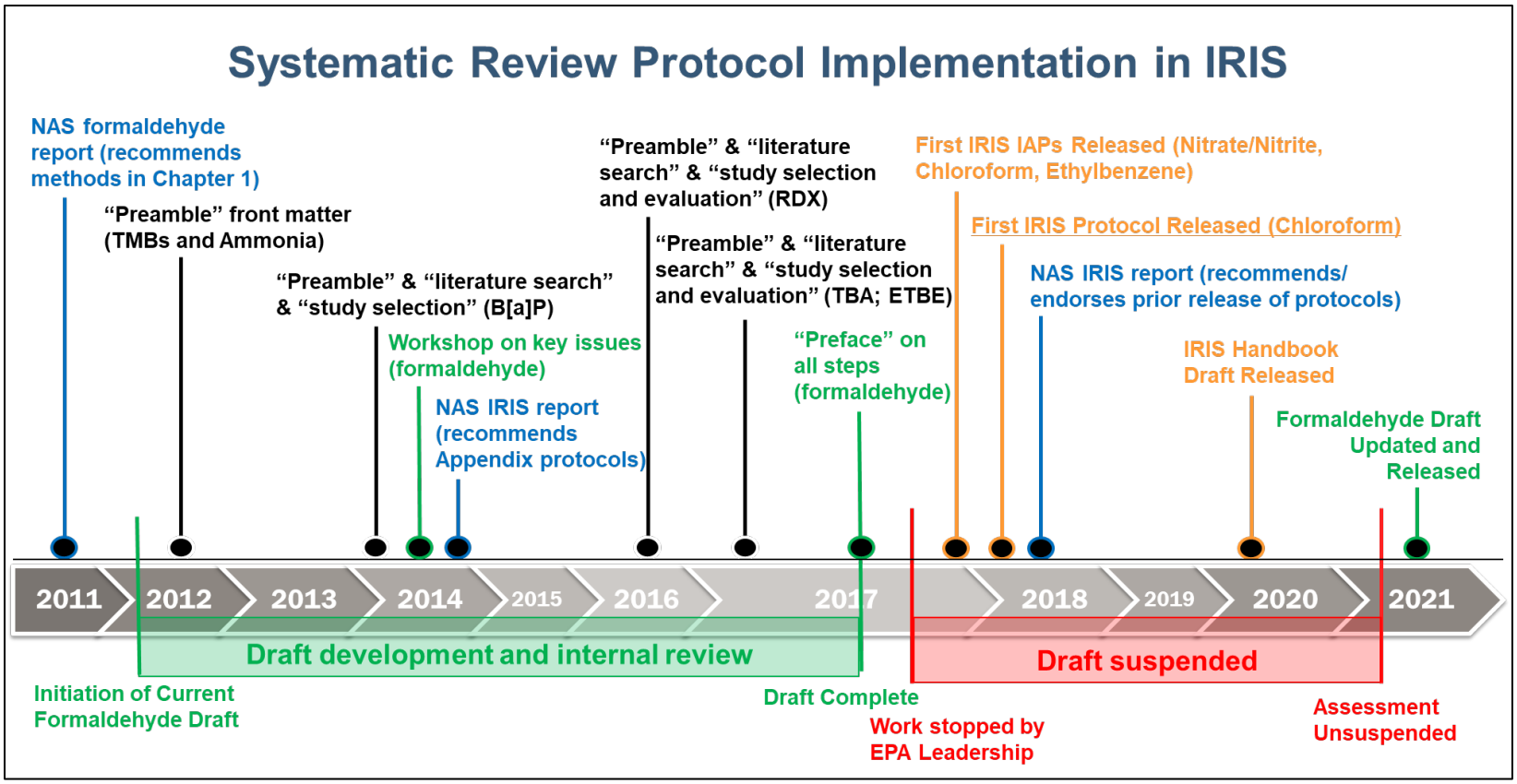


*Figure 2. Illustrative Timeline of Protocol Implementation in the IRIS Program*

Details and links to materials provided in the Addendum. Abbreviations: TMBs, Trimethylbenzenes; B[a]P, Benzo[a]pyrene; RDX, Hexahydro-1,3,5-trinitro-1,3,5-triazine; TBA, tert-butyl alcohol; ETBE, Ethyl tertiary butyl ether; IAP, IRIS assessment plan (scoping and problem formulation document).

a. Specifically, can EPA provide protocols for the multiple reviews of the various non-cancer outcomes and cancer types, encompassing the human and animal evidence and covering the eight steps outlined in the figure?

EPA Response to Question 1a:

As noted above, descriptions of the methods for evaluating individual studies and synthesizing the evidence for each non-cancer outcome and cancer type are detailed in the assessment "Preface on Assessment Methods and Organization" in the Toxicological Review, with the additional outcome-specific considerations for study evaluation outlined in Appendix A.5. In addition, the synthesis of evidence on each outcome includes a section, "methodological issues considered in the evaluation of studies" that elaborates on the considerations for interpreting the evidence from the available studies on each outcome. This response will focus on the outcome-specific methods and considerations applied in evaluating and synthesizing the human and animal evidence. Table 1 below provides a more granular mapping of the eight steps outlined in Figure 1 to where those methods and documentation for evaluating the apical human and animal evidence on each noncancer outcome and cancer type can be found in the draft assessment materials.[5] Notably, these methods draw heavily from existing outcome-specific recommendations in EPA guidelines and other authoritative sources (e.g., WHO guidance).

In addition to the locations of information on the methodological approaches for each of the eight steps in Table 1, the bullets below highlight specific areas of emphasis and the guidelines or other materials that were used in evaluating and synthesizing the available evidence on each health outcome. As noted in response to Question 2, the considerations and criteria for reviewing the various health effects stemmed directly from guidance provided in the NAS (2011) report (e.g., compare the recommendations on p. 98-99 for evaluating animal studies of neurotoxicity with the criteria bulleted below). However, this list does not focus on responsiveness to the NAS (2011) report as that is highlighted elsewhere (see Appendix D of the draft IRIS formaldehyde assessment materials). Primarily, the details below are from the "Study evaluation" subsections within Appendix A.5 and the "Methodological issues…" subsections of each synthesis section in Sections 1.2 and 1.3 of the Toxicological Review.

[5] The evaluation and synthesis of the mechanistic evidence relevant to each health outcome is not detailed here (for details, please see the "Preface" and the introductory text to the MOA sections for each noncancer effect and cancer type grouping in Sections 1.2 and 1.3 of the Toxicological review, and in the Appendices Sections A.4, A.5.1, and A.5.4-A.5.9 in particular).

- Sensory irritation (Sections 1.2.1, A.5.2)
    - Human evidence[6]: The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. For this outcome, they emphasized exposures in the referent group, blinding, exposure assessment, and the time frame separating exposure and outcome given the acute nature of irritant responses.
    - Animal evidence: not formally evaluated, but well-established (Appendix A.3).

- Pulmonary function (Sections 1.2.2, A.5.3)
    - Human evidence: EPA used guidelines from the American Thoracic Society to interpret the appropriateness of the pulmonary function measures used in studies as well as the biological significance of small changes in these functions. Of particular importance were the exposure assessment, and the outcome assessment with respect to the adjustments of lung function by race/ethnicity, gender, age, and height and evaluations of potential healthy worker or survivor effects (lead time bias) were key considerations for these outcomes.
    - Animal evidence: not formally evaluated (few studies available; difficult to measure accurately in experimental animals; and many human studies available).

- Immune-mediated conditions, focusing on allergies and asthma (Sections 1.2.3, A.5.4)
    - Human evidence: EPA developed and applied criteria for evaluating these effects based on consultation with panels of five international experts on allergy and five international experts on asthma (one expert served on both consultation panels). The consultations were conducted through a contract mechanism. Specifically, the contractor: identified and recruited the panelists based on their expertise; provided to the panelists a set of questions developed by EPA (e.g., on the definitions, interpretation, and potential grouping of different outcomes reported in the available studies; on the validity, appropriateness, and interpretability of the instruments used to measure the outcomes in different populations and lifestages); collected written responses to EPA's questions from the panelists; and facilitated discussions between EPA and the experts via teleconference. These consultations resulted in criteria focusing on the types of measures

[6] The evaluation of human studies across all health effects applied standard considerations within the epidemiological discipline, the bullets here capture points of emphasis particular to each health effect.

reported in the studies and their interpretation during evidence synthesis, as well as use of validated questionnaires and relevant exposure periods.
    - Animal evidence: Ultimately, based on consideration of the International Programme on Chemical Safety (IPCS) guidelines, these data were evaluated as mechanistic information.

- Respiratory Tract Pathology (Sections 1.2.4, A.5.5)
    - Human evidence: The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized use of standard methods for histopathological examinations and study handling of potential confounding.
    - Animal evidence: The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment, although outcome-specific criteria were informed by methods applied in good laboratory practice (GLP) studies on these outcomes (e.g., preferred sample size, blinding procedures, and histopathological sampling). In addition to these areas and because of the robust database, the evidence synthesis also emphasized consideration of study duration and tested formaldehyde levels.

- Respiratory Tract Cancers (Sections 1.2.5, A.5.9)
    - Human evidence: EPA developed and applied criteria for evaluating these effects. The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized exposure measurement error in general and with respect to cancer latency, confounding, and study sensitivity for the rare cancer outcomes.
    - Animal evidence: The approach applied was similar to that for respiratory tract pathology studies in experimental animals, incorporating and applying recommendations on study design (e.g., sample size; study duration; consideration of historical controls) and evidence interpretation described in the 2005 EPA cancer guidelines (HERO ID 6324329).

- Nervous System Effects (Sections 1.3.1, A.5.7)
    - Human evidence: EPA developed and applied criteria for evaluating these effects. The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized consideration of irritant contributions to behavioral measures and the influence of potential confounding, quality of the exposure measures, and numbers of exposed cases for nervous system disease measures.

- Animal evidence: EPA developed and applied criteria for evaluating these effects. The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized considering the irritant and stress-inducing properties of formaldehyde on tested behaviors, as well as the duration and periodicity of formaldehyde exposures and sufficiency of endpoint evaluations in comparison with EPA guidelines. In addition to systematically evaluating these factors and basing judgments on application of the 1998 EPA neurotoxicity guidelines (HERO ID 30021), an additional major factor influential to the interpretation of all systemic effects was the appropriateness of the test article used.

- Developmental and Reproductive Toxicity (Sections 1.3.2, A.5.8)
  - Human evidence: EPA developed and applied criteria for evaluating these effects. The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized participant selection with respect to pregnancy gravidity, outcome ascertainment, potential recall bias and confounding.
  - Animal evidence: These evaluations generally follow the EPA guidelines for evaluating reproductive toxicity (HERO ID 7581; 1996) and developmental toxicity (HERO ID 732120; 1991), including evaluating potential maternal toxicity, appropriateness of the timing of exposure (e.g., critical windows), and comparisons of the study methods (including randomization and blinding) against established standards for commonly evaluated reproductive and developmental endpoints. In addition, as with other systemic effects, the test article used was a major consideration.

- Lymphohematopoietic Cancers (Sections 1.3.3, A.5.9)
  - Human evidence: EPA developed and applied criteria for evaluating these effects. The evaluations were not based on protocols apart from the methods presented in the draft IRIS formaldehyde assessment. They emphasized exposure measurement error in general and with respect to cancer latency and study sensitivity for the rare cancer outcomes.
  - Animal evidence: In general, the approach applied was parallel to that used for evaluating respiratory tract pathology and respiratory tract cancers, including a focus on sample size, blinding, and adequacy of sampling, but as with other systemic effects a major emphasis was on the appropriateness of the test article.

*Table 1. Mapping to Description of Methods for Each of the 8 Steps in Figure 1*

| **Health Effect[1]** | **1. Reference search and retrieval** | **2. Reference Screening** | **3. Outcome-Specific Evaluation Criteria** | **4. Study Evaluation Tables** | **5. Syntheses of Results** | **6. Evidence Synthesis Judgments** | **7. Evidence Integration Judgments** | **8. Study Selection for Dose-Response** |
|---|---|---|---|---|---|---|---|---|
| General | Preface<br>A.5.1, page A-231 | Preface | N/A, see below | A.5.1, pages A-232-239[2] | Preface | Preface | Preface | Preface |
| Sensory irritation | Section: A.5.2 | Section: A.5.2 | Pages: A-263-264 & Table A-33 | Tables A-34 to A-39 | Section 1.2.1 | Pages: 1-32-34 & Table 1-4 | Sections 1.2.1, 1.4.2 | Section 2.1.2 |
| Pulmonary function | Section: A.5.3 | Section: A.5.3 | Pages A-301 & Table A-43 | Tables A-44 to A-45 | Section 1.2.2 | Pages: 1-75-77 & Table 1-11 | Sections 1.2.2, 1.4.2 | Section 2.1.2 |
| Immune-mediated conditions | Section: A.5.4 | Section: A.5.4 | Pages: A-342-349 & Table A-50 | Tables A-51 to A-52 | Section 1.2.3 | Pages: 1-146-149 & Table 1-24 | Sections 1.2.3, 1.4.2 | Section 2.1.2 |
| Respiratory tract pathology | Section: A.5.5 | Section: A.5.5 | Pages A-594 & Table A-57 | Tables A-58 to A-61 | Section 1.2.4 | Pages: 1-196-196 & Table 1-30 | Sections 1.2.4, 1.4.2 | Section 2.1.2 |
| Nervous system effects | Section: A.5.7 | Section: A.5.7 | Pages A-592-593 & Table A-84 | Tables A-85 to A-87 | Section 1.3.1 | Pages: 1-377-382 & Table 1-50 | Sections 1.3.1, 1.4.2 | Section 2.1.2 |
| Developmental & reproductive toxicity | Section: A.5.8 | Section: A.5.8 | Pages A-636-639 & Table A-91 | Tables A-92 to A-93 | Section 1.3.2 | Pages: 1-430-434 & Table 1-58 | Sections 1.3.2, 1.4.2 | Section 2.1.2 |

| Health Effect[1] | 1. Reference search and retrieval | 2. Reference Screening | 3. Outcome-Specific Evaluation Criteria | 4. Study Evaluation Tables | 5. Syntheses of Results | 6. Evidence Synthesis Judgments | 7. Evidence Integration Judgments | 8. Study Selection for Dose-Response |
|---|---|---|---|---|---|---|---|---|
| Upper respiratory tract cancers | Section: A.5.9 | Section: A.5.9 | Pages A-676-687 & Table A-105 | Tables A-106 to A-108 | Section 1.2.5 | Pages: 1-336-341 & Table 1-43 | Section 1.4.3 | Section 2.2.1 |
| Lympho-hematopoietic cancers | Section: A.5.9 | Section: A.5.9 | Pages A-676-687 & Table A-105 | Tables A-106 to A-108 | Section 1.3.3 | Pages: 1-541-544 & Table 1-67 | Section 1.4.3 | Section 2.2.2 |

[1]As described in the Preface and Appendix A.5.1, approaches for select literature searches (e.g., see Appendix A.5.6), study evaluations (e.g., see Appendix A.4, A.5.1 and A.5.6), and evidence synthesis considerations (e.g., see Section 1.2.5) for mechanistic evidence relevant to one or more of the different health outcomes in the above Table were also provided.

[2]This description of methods includes standardized approaches for evaluating individual observational epidemiology studies, controlled human exposure studies, experimental animal studies, mechanistic studies, as well as evaluating exposure assessment across epidemiology studies (the health effect-specific considerations for rating exposure in epidemiology studies are outlined in Appendix A.52 through A.5.9; see Table A-43 for example) and exposure quality in controlled inhalation exposure studies in humans or animals (documented for individual formaldehyde studies in pages A-240 to A-260 and described in more detail in the following publication: https://doi.org/10.1016/j.toxlet.2019.05.011).

b) Can EPA provide more detailed information about how the study quality determinations were made, for example, that harmonizes the general and the endpoint-specific study quality information provided in Appendix A.5?

EPA Response to Question 1b:

As described in the Preface and Appendix A.5.1, the evaluation of individual studies followed a structured process conducted on an outcome-specific basis. Study evaluation was conducted for all studies meeting PECO criteria, independent of the direction, magnitude, or statistical significance of the study's results. Determinations of overall confidence (*high*, *medium*, or *low* confidence, or *not informative)* are based on domain judgments of considerations such as population selection, exposure quality, potential for confounding, etc. (see bullets below for general criteria for judging confidence from Appendix Tables A-28 and A-29). The direction of influence on the results of any identified sources of potential bias or insensitivity is expressed as a qualitative narrative expert judgment, i.e., no algorithmic or quantitative scoring approach is used. The rationale for the overall confidence judgment is articulated in the Appendix study evaluation tables (Appendix A.5), with salient aspects brought into the evidence synthesis sections (sections 1.2 and 1.3). High confidence studies generally had no notable limitations (no notable downgrades in domain-specific judgments), medium confidence studies were generally well-conducted but had issues that might introduce a minor amount of uncertainty (some minor downgrades within one or a few domains), and low confidence studies had significant deficiencies that affect interpretability (major downgrade in one or more domains). Thus, judgments of confidence were harmonized across the three disciplines to reflect a common interpretation of the impact of limitations identified during evaluation on the reliability or interpretability of the reported results.

The approaches were harmonized by discipline area, so the "study attribute" or "experimental feature" categories (i.e., termed "domains" here) considered did vary across observational epidemiology, controlled human exposure, and experimental animal studies.[7] Within each discipline, the domains evaluated were consistent across health outcomes, although the influence of ratings for specific domain was sometimes more impactful on the confidence judgment for one health effect as compared to another when considering the potential influence of the identified deficiency on the outcome. Appendix A.5.1 describes the general considerations for evaluating each domain across the three discipline areas (see below), with focused or augmented health effect-specific considerations applied to the rating criteria within each domain outlined

[7] The evaluation of individual mechanistic studies on select topic areas aligned with simplified domains for these aforementioned three disciplines, as described in Appendix A.4 and A.5.

in Appendix A.5.2-A.5.9 (see examples for epidemiology and experimental animal studies in Tables 2 and 3 below). The general domain-specific criteria were developed *a priori* (i.e., before reviewing the studies). The health outcome-specific refinements to the criteria evaluated within domains were based on the available endpoints and study designs within the discipline- and health effect-specific evidence base. In some instances, adjustments to the criteria were necessary after the study evaluations began; any such adjustments were then applied uniformly across all studies within that discipline and health effect. These adjustments were due to identification of an important feature of the evidence only as the individual studies were being reviewed (e.g., evaluations of confounding across occupational studies considered the potential co-exposures in the workplace, which were sometimes revealed within an individual study and had to then be considered uniformly across studies within that same occupational setting).

- Epidemiology study general considerations for evaluating each domain (p. A-233):

*Population Selection (SB)*: Recruitment, selection into study, and participation independent of exposure status and reported in sufficient detail to understand how subjects were identified and selected.

*Information Bias (IB)*: Validated instrument for data collection described or citation provided. Outcome ascertainment conducted without knowledge of exposure status. Timing of exposure assessment appropriate for observation of outcomes. Information provided on the distribution and range of exposure with adequate contrast between high and low exposure.

*Potential for confounding (Cf)*: Important potential confounders addressed in study design or analysis. Potential confounding by relevant co-exposures addressed.

*Analysis (Oth: "other features of design or analysis")*: Appropriateness of analytic approach given design and data collected; consideration of alternate explanations for findings; presentation of quantitative results.

*Other considerations not otherwise evaluated (Oth)*: Sensitivity of study (exposure levels, exposure contrast, duration of follow-up, sensitivity of outcome ascertainment).

- Controlled human exposure study general considerations for evaluating each domain (p. A-233):

As noted in the Preface, "a process incorporating aspects of the evaluation approaches used for epidemiological studies and experimental animal studies was used to evaluate controlled exposure studies in humans." Thus, in addition to considering bias domains evaluated for

epidemiology studies as appropriate for the study design (noting that human exposure studies were able to evaluate symptoms in a controlled environment and therefore, the exposure-response relationship was more precise, and potential confounders were of less concern), the following considerations specific to experimental studies (see also experimental animal study domains below) were evaluated: randomization of exposure assignments, blinding of subjects and investigators, number of individuals evaluated, and exposure quality (separately documented in Appendix A.5.1).

- Experimental animal study general considerations for evaluating each domain (p. A-234):

*Exposure Quality:* Given the importance of the inhalation exposure paradigms used across the available experimental animal studies, detailed evaluations of exposure quality were separately performed for each study (see Appendix A.5.1, Exposure Quality Evaluation: Animal Toxicology and Controlled Human Exposure Studies).

*Test Animals:* The species, sex, strain, and age are considered appropriate and sensitive for testing the endpoint(s); sample size provides reasonable power to assess the endpoint(s); overt systemic toxicity is absent or not expected at the tested concentrations, or it is appropriately accounted for. Groups appear to be adequately matched at the onset of the experiment.

*Study Design:* The study design is appropriate and informative for evaluating the endpoint(s), including a sufficient exposure duration and/or appropriate timing of endpoint evaluations to allow for sensitive detection of the effect(s) of interest, and a lack of additional variables introduced over the course of the study that would be expected to modify the endpoint(s).

*Endpoint Evaluation:* The protocols used to assess the endpoint(s) are sensitive (able to detect subtle changes in the health outcome of interest), complete (include the appropriate protocol controls), discriminating (specific for the health outcome in question), and biologically sound (note: this applies to evaluations of novel or unproven methods regarding their ability to detect the changes in the endpoints of interest). The potential for experimenter bias is minimized.

*Data Considerations and Statistical Analysis:* Data for all endpoints evaluated in the study are presented with sufficient detail (e.g., variability is included) and in the preferred form (e.g., arbitrary cut-offs were not applied to continuous data). Statistical methods and the group comparisons analyzed appear to be completely reported, appropriate, and discerning (note: when inappropriate statistical methods appear to have been used, EPA sometimes performed additional comparisons, as documented in evidence tables).

Harmonizing these evaluations by using consistent definitions of confidence, evaluating the same sources of bias and insensitivity within each discipline using conserved domains across health outcomes, and using the same general criteria for evaluating those domains within each discipline area across health outcomes facilitates consistent application. In addition, several workflow processes were implemented to further promote consistency. The approach for study evaluation involved an initial discussion amongst topic-specific experts to develop health outcome- and discipline-specific criteria to apply during the evaluation. Although this was typically internal to EPA, topic-specific expertise on epidemiology studies of immune-mediated conditions was sparse and external experts were consulted (documented in Appendix A.5.4). The evaluation of each study involved an initial review by a primary topic-specific expert and a secondary review by a second expert who also reviewed the extracted domain-specific details for accuracy (the secondary reviewer was not blinded to the primary review). Disagreements across the two reviewers were addressed through discussion and consultation of the source paper, with a third reviewer added to address any disagreements that could not be resolved. Only the final judgments were documented, and this documentation focused primarily on any identified limitations. This approach is consistent with methods used within the IRIS Program at the time (~2012–2015), prior to the development and adoption of dedicated software tools that allowed for more independent study evaluations and simplified transparent documentation. In addition to the reviews of individual studies by the two topic-specific experts, discipline-specific experts (e.g., epidemiologists) on the team met to discuss judgments on studies within their domain across health effects to ensure consistency in the judgments, their application within the relevant evidence synthesis sections, and presentation in tables and figures. The formaldehyde team and other contributors to earlier systematic review steps (i.e., pre-evidence synthesis) during the timeframe of 2012-2015 included approximately 4-5 epidemiologists and 6-9 toxicologists. Each health effect-specific section was also formally reviewed internally by topic-specific workgroups (each consisting of ~5-10 experts each, typically with a primary and sometimes secondary reviewer that raised comments for discussion with the broader WG). Disciplinary workgroups within the IRIS Program at the time included: Epidemiology; Inhalation; General Toxicology (hepatic, renal, etc.), Cancer, and Immune; Developmental, Reproductive, and Nervous System; Pharmacokinetic; Quantitative Methods; and Toxicity Pathways Workgroups.

Finally, the draft assessment applied consistency in the presentation of these evaluations across health effects to more transparently convey the methods applied and decisions made on individual studies within and across health effects. This is evident in the inclusion of similar sub-sections outlining health outcome-specific considerations for evaluating individual studies (e.g., the “Methodological considerations…” sections within each evidence synthesis section; the “Study Evaluations” sections outlining health effect-specific considerations on each discipline-specific set of studies within Appendices A.5.2-A.5.9). This is also apparent in the presentation across health effects

of domain-specific and overall judgments by discipline in the study evaluation tables in the Appendices. Specifically, domain-specific decisions for observational epidemiology studies were documented using consistent graphics to depict the rating and direction of bias (note: controlled human exposure studies were documented using text only and did not apply these graphics to depict potential bias since they were controlled studies), while documentation of experimental animal studies used consistent symbols and shading ('++' = robust/no impactful limitations, '+' = adequate/potential issues identified but not expected to have a substantial impact on results, and shaded cells = poor/significant limitation identified that is expected to substantially impact results). This consistent documentation is likewise conserved across evidence tables and figures in the Toxicological Review (e.g., shading of low confidence experimental animal studies across health effects). This effort to provide consistency in presentation to facilitate comparison of the assessment judgments across the disparate noncancer and cancer health effects section was similarly applied to all other aspects of documenting the assessment methods and results (literature search and screening; evidence synthesis and integration; etc.).

*Table 2. Animal Study Evaluation Examples of Health Effect-Specific Considerations Most Impactful to Domain Interpretations*

| | **Exposure Quality[1]** | **Test Subjects** | **Study Design** | **Endpoint Evaluation** | **Data Considerations** |
|---|---|---|---|---|---|
| Sensory irritation | N/A, systematic review focused on human studies, although mechanistic understanding was impactful | | | | |
| Pulmonary function | N/A, systematic review focused on human studies, with some indirect mechanistic support for judgments | | | | |
| Immune conditions | N/A, as noted in Appendix A.5.4, experimental animal studies were specifically evaluated as mechanistic information (documented in Appendix A.5.6) | | | | |
| Respiratory tract Pathology (A.5.5; Table A-59 header & preceding text) | Inhalation exposure administration methods were evaluated independent of health outcome (see Appendix A.5.1) with the exception of test article[2] | Large effect on survival or, given strong database, N<10/group = downgrade | Given slow-developing lesions, short exposure duration = downgrade | Low sampling of target tissues (URT; nose) = downgrade | Not separately reporting incidence for different lesions = downgrade |
| Developmental & reproductive toxicity (A.5.8; Table A-93 header & preceding text) | | Overt toxicity or inadequate allocation = downgrade | Lack of control for potential confounding = downgrade | Insensitive or incomplete testing methods = downgrade | Lack of accounting for potential litter effects or selective reporting = downgrade |
| Nervous system effects (A.5.7; Table A-86 header & preceding text) | | Lack of testing both sexes or overt toxicity = downgrade | Given irritant effects, short latency between exposure and behavior testing = downgrade | Lack of blinding or sampling bias, or insensitive methods = downgrade | Lack of accounting for potential litter effects or selective reporting = downgrade |
| Cancer (A.5.9; Table A-107 header & preceding text) | | Unaddressed decreases in survival or, given strong database, N<20/ group = downgrade | Given slow-developing lesions, short exposure duration = downgrade | Low sampling of target tissues or lack of blinding = downgrade | Inadequate reporting of lesions (e.g., location; incidence) or mortality = downgrade |

[1]This table focuses on apical experimental animal data, although exposure quality evaluations were also performed for studies of controlled human exposure. Additionally, some of these same considerations were applied to evaluating mechanistic studies (e.g., see Appendix A.4; A.5.6).

[2]Given the concern that inhaled methanol (unlike formaldehyde) can be systemically distributed, the evaluation of the test article characterization and controls domain was more impactful to exposure quality judgments for systemic effects than respiratory (POE) effects (assigning a rating of *poor* versus *adequate*, respectively).

*Table 3. Epidemiology Study Evaluation Examples of Health Effect-Specific Considerations Resulting in Downgrades in a Domain Rating*

| | **Consideration of participant selection and comparability** | **Exposure measure and range** | **Outcome measure** | **Consideration of likely confounding** | **Analysis and completeness of results** | **Size** |
|---|---|---|---|---|---|---|
| Sensory irritation (A.5.2; Table A-34 & preceding text) | No comparison group;<br>Referent group also exposed;<br>Lack of blinding / recall bias;<br>Low participation rate;<br>Healthy worker and survivor effect | Limited exposure assessment;<br>Short or minimal exposure | Time frame separating exposure and outcome | If co-exposures to other identified causes of outcome;<br>Failure to adjust for key confounders | Inadequate reporting | Limited number of case events |
| Pulmonary function (A.5.3; Table A-44 & preceding text) | No comparison group;<br>Low participation rate;<br>Healthy worker and survivor effect | Limited exposure assessment;<br>Short or minimal exposure | Time frame separating exposure and outcome | If co-exposures to other identified causes of outcome;<br>Failure to adjust for key confounders | Inadequate reporting | Limited number of case events |
| Immune-mediated conditions (A.5.4 Table A-51 and preceding text) | Referent group also exposed;<br>Low participation rate | Short or minimal exposure;<br>Sampling period not reported or uncertain | Specificity of outcome | If co-exposures to other identified causes of outcome;<br>Failure to adjust for key confounders | Inadequate reporting | Limited number of case events |

| | Consideration of participant selection and comparability | Exposure measure and range | Outcome measure | Consideration of likely confounding | Analysis and completeness of results | Size |
|---|---|---|---|---|---|---|
| Respiratory tract Pathology (A.5.5; Table A-58 header & preceding text) | Referent group also exposed;<br>Healthy worker and survivor effect | Short or minimal exposure;<br>Sampling period not reported or uncertain | Specificity of outcome | If co-exposures to other identified causes of outcome;<br>Failure to adjust for key confounders | Inadequate reporting | Limited number of case events |
| Nervous system effects (A.5.7; Table A-84 header & preceding text) | No comparison group;<br>Referent group also exposed | Uncertainty in exposure assessment | Specificity of outcome | If co-exposures to other identified causes of outcome;<br>Limited covariate data | Not accounting for latency using lagged exposure | Limited number of case events |
| Developmental & reproductive toxicity (A.5.8; Table A-93 header & preceding text) | Participation related to exposure or outcome;<br>Loss to follow-up | Uncertainty in exposure assessment;<br>Sampling period not reported or uncertain | Specificity of outcome | Failure to adjust for key confounders;<br>Adjustment for previous pregnancy loss may introduce bias | Inadequate reporting | Limited number of case events |
| Cancer (A.5.9; Table A-107 header & preceding text) | Participation %;<br>Participation related to exposure or outcome;<br>Loss to follow-up;<br>Use of next-of-kin;<br>Healthy worker effect;<br>Prevalent cases vs. incident | Reliability and sensitivity of exposure measure;<br>Exposure assigned by industrial setting or single job;<br>Minimal exposure | Specificity of outcome | If co-exposure to other identified causes of the same cancer | Inappropriate analyses with respect to design;<br>Not accounting for cancer latency using lagged exposure | Limited number of case events;<br>Short length of follow-up |

c) How was consistency in approach assured across multiple working groups, particularly within each endpoint?

EPA Response to Question 1c:

Consistency was maintained through the use of multiple layers of review within each endpoint and across all endpoints. Much of this process is described in the answer above in the context of study evaluation, but to reiterate and expand on some points:

- To promote consistency in approaches across the multidisciplinary formaldehyde assessment team, the chemical managers worked with team members to develop standardized templates for documenting and presenting the various decision steps common across sections (e.g., synthesis and integration section outline templates; evidence table, evidence integration summary table, and study evaluation table templates with conserved formatting and use of graphics, symbols, and shading by discipline; literature flow diagram templates). For instance, reviewers performing study evaluations were provided with template (empty) versions of the study evaluation tables[8] in Appendix A.5 alongside the associated discipline-specific general instructions and health effect-specific criteria (e.g., from the "Study evaluations" subsections, with criteria also included in the table headings for animal studies) for answering each domain.

- Outcome-specific criteria informing the evaluation of the different study evaluation domains were discussed amongst topic-specific experts, in some instances including experts outside of EPA.

- Each study was evaluated by a primary and a secondary topic-specific expert reviewer who reached agreement on the confidence conclusions. The evaluations placed an emphasis on exposure considerations; for controlled inhalation exposure studies, exposure quality was separately evaluated by inhalation toxicology experts.

- Multiple internal reviews of the syntheses of each endpoint by other authors within the same discipline (e.g., epidemiology) ensured that the common application of the approaches to study evaluation and synthesis were maintained.

[8] An example for observational epidemiological studies on noncancer (pulmonary function) is Table A-44 and cancer is Table A-105; an example for experimental animal studies of nervous system effects is Table A-86; and an example for controlled human exposure studies (sensory irritation) is Table A-36.

- Review by the chemical managers across all endpoints ensured that the common application of the approaches to study evaluation and synthesis, including presentation of those decisions and the associated documentation, were maintained.

- Each evidence synthesis section and the underlying analyses were evaluated by members of one or more IRIS program disciplinary workgroups who were not team members on the assessment (details in response to Question 1b).

- As with all IRIS assessments developed within the timeframe of 2014-2017, the draft IRIS formaldehyde assessment underwent an executive level review by a subset of senior managers.

- IRIS assessments, including this draft IRIS formaldehyde assessment, undergo Agency and Interagency reviews before being released for public comment and subsequent peer review. These multiple review steps help to ensure consistency in the approaches applied and the decisions made in the final, posted assessment.

2) Can EPA describe the general guidance that was behind the systematic review protocols over time, from 2011 forward? How did they respond to guidance (*i.e.*, from prior peer reviews)?

EPA Response to Question 2:

The primary guidance used to develop the current draft IRIS assessment of formaldehyde was the 2011 NAS formaldehyde report. This is worth emphasizing. The draft IRIS formaldehyde assessment is unique in that a roadmap of methods and recommendations specifying how EPA should go about developing a more robust approach to the draft assessment were outlined in detail in the 2011 NAS formaldehyde report. Thus, the methods and approaches recommended in the 2011 report were core components of the methods applied to develop the current draft (see Appendix D of the draft IRIS formaldehyde assessment materials and Table 4 below). Thus, in many ways, the 2011 NAS report provided an early and publicly available description of the methods that would be used to develop the current draft. This provides some measure of fulfilling the functional role of a protocol in current practice with respect to prior release of pre-specified methods used to conduct the assessment.

The methods utilized in the assessment were also informed by engagements the IRIS Program organized between 2012 and 2017 to develop its systematic review methods (see the Addendum, particularly Table A.2). The development of the IRIS Handbook began around 2012, with the methods being developed and evolving in parallel with the drafting of the draft IRIS formaldehyde report. The methods developed and optimized for inclusion in the IRIS Handbook were piloted in the drafting of the draft IRIS formaldehyde assessment. In fact, as previously noted, several authors of the draft IRIS formaldehyde assessment were also core authors drafting the IRIS Handbook. Thus, the underlying methods in the evolving and posted IRIS Handbook and the current draft assessment are fundamentally the same. The methods in the IRIS Handbook were directly informed by and responsive to guidance and recommendations received across an array of inputs to the IRIS Program between 2012 and 2020 (when the draft IRIS Handbook was publicly released, after the assessment was suspended), including NAS reviews in 2011, 2014, and 2018; public workshops on systematic review; testing of different methods applied during the various steps of systematic review (as evidenced by some publications on these topics coauthored by members of the formaldehyde assessment team); and peer review feedback on draft assessments incorporating evolving aspects of systematic review.

The similarities between the evolving Handbook methods and the methods used to develop the draft IRIS formaldehyde assessment draft are evident when looking at some of the materials made public along the way to releasing the completed Handbook. For example, EPA publicly released in 2013 materials to inform the NAS 2014 review of the

evolving IRIS Handbook methods; the materials included updates on the status of implementing the prior NAS recommendations (U.S. EPA, 2013, HERO ID 1511259) and chemical-specific examples (U.S. EPA, 2013, HERO ID 1511260). The status document included a draft version of the IRIS Handbook in Appendix F. Close parallels between the 2014 methods presented in Appendix F and the methods and documentation used in the draft IRIS formaldehyde assessment are apparent, particularly for the early and middle steps of the 8-step process laid out in Figure 1. A couple specific examples are highlighted. Figure 3 below on documenting literature search and screening decisions includes Figure F-1 from Appendix F (U.S. EPA, 2013, HERO ID 1511259) shown at left and a literature tree from the draft IRIS formaldehyde assessment at right (Appendix A.5). Likewise, Table 5 presents the domains used for evaluating epidemiology (Table F-6) and experimental animal (Table F-7) studies in Appendix F (U.S. EPA, 2013, HERO ID 1511259) alongside the domains evaluated in the draft IRIS formaldehyde assessment (see Preface and Appendix A.5), both of which build from recommendations made in the NAS (2011) report (also shown in Table 5). Further, Appendix F (U.S. EPA, 2013, HERO ID 1511259) includes descriptions of how these domains are applied, including specific questions and considerations that largely match the content in the draft IRIS formaldehyde assessment. As is shown in Figures F-8 and F-9 of Appendix F (U.S. EPA, 2013, HERO ID 1511259), the study evaluation documentation tables for epidemiology studies, controlled inhalation exposures, and animal studies are nearly identical (as they reflect examples from the draft IRIS formaldehyde assessment at that time). Likewise, templates for data extraction and reporting results (evidence tables) in Appendix F (U.S. EPA, 2013, HERO ID 1511259) mirror approaches used in the development of the draft IRIS formaldehyde assessment. The materials shared with the 2014 NAS committee present some considerations for evidence synthesis and integration, and selection of studies and datasets for dose-response analysis, but the methods had not yet evolved to the point of transparency and rigor used in developing the draft IRIS formaldehyde assessment. The IRIS Program engaged on these topics, particularly evidence synthesis and integration, with experts after 2014, including receiving specific peer feedback on them from the 2018 NASEM committee (see link above; the public EPA presentation is in Appendix C of the NASEM report). This resulted in refinements to the presentation of some of these aspects in the current draft IRIS formaldehyde assessment (e.g., standardized documentation of evidence integration decisions).

As evolving systematic review methods were applied within peer reviewed assessments such as TMBs and B[a]P, the feedback received during the review of those documents directly informed revisions to the methods included in both the emerging IRIS Handbook and the draft IRIS formaldehyde assessment (links provided in the Addendum). In addition, as depicted in Table A.2 in the Addendum and the slide deck previously presented to this committee at the October 2022 meeting, the IRIS Program engaged with the EPA Science Advisory Board on the preliminary assessment development methods and consulted directly with multiple international, federal, state, and academic panels of systematic review experts, including the 2014 NASEM panel, on

the methods the IRIS Program was developing for inclusion within the IRIS Handbook. And, more specifically focusing on the draft IRIS formaldehyde assessment, to inform the methods applied to evaluate key science issues associated with interpreting some of the potential health effects of inhaled formaldehyde, the IRIS Program hosted a public workshop on several assessment key science issues (see Addendum for details) and consulted with a group of international experts on asthma and allergic diseases (see Appendix A.5.4), both in 2014.

These various engagements, reviews, and inputs to the IRIS Program more broadly, and on the draft IRIS formaldehyde assessment specifically, directly shaped the evolving direction of the literature search and screening, study evaluation, evidence synthesis and integration, and dose-response methods used in drafting the IRIS formaldehyde assessment, as well as the systematic evidence mapping approach used to update the assessment after it was unsuspended in 2021.

*Table 4. Abbreviated Crosswalk of NAS (2011) Report Recommendations to 2022 IRIS Formaldehyde Draft Assessment*

| **NAS 2011 "critical" recommendations[a]** | **In 2022 draft (see also Appendix D of the draft IRIS assessment)** |
|---|---|
| "First, rigorous editing is needed to reduce the volume of the text substantially and address the redundancies and inconsistencies; reducing the text could greatly enhance the clarity of the document." | • Does not include study-by -study descriptions<br>• Synthesis text shortened and reorganized around health effect; tables and figures made more illustrative<br>• Assessment Overview document shows rigorously edited summary, although this level of detail is generally insufficient for some EPA purposes. |
| "Second, Chapter 1 of the draft assessment needs to discuss more fully the methods of the assessment. The committee is recommending not the addition of long descriptions of EPA guidelines, but rather clear concise statements of criteria used to exclude, include, and advance studies for derivation of the RfCs and unit risk estimates." | • Preface on assessment methods (and supporting Appendices) describes all the requested criteria.<br>• Note that NAS (2011) did not recommend separately released protocol, but inclusion of methods in "Chapter 1 of the draft assessment" |
| "Third, standardized evidence tables that provide the methods and results of each study are needed for all health outcomes; if appropriate tables were used, long descriptions of the studies could be moved to an appendix or deleted." | • Fully embraced, as evidenced by summary evidence tables for each health outcome and not including long descriptions of each study. |
| "Fourth, all critical studies need to be thoroughly evaluated for strengths and weaknesses by using uniform approaches; the findings of these evaluations could be summarized in tables to ensure transparency." | • Uniform approaches, by discipline, were applied to evaluate study strengths and weaknesses. These are detailed in the assessment Preface, with additional considerations provided in the health-effect specific Appendices.<br>• As recommended, the findings of these evaluations are summarized in study evaluation tables by health effect and discipline in the Appendices. |
| "Fifth, the rationales for selection of studies that are used to calculate RfCs and unit risks need to be articulated clearly." | • General considerations for selecting studies for toxicity value derivation are presented in the Preface, Table X.<br>• Rationales for selecting studies to derive PODs and candidate values for consideration in RfC development are presented in tables by health effect (see |

| NAS 2011 "critical" recommendations[a] | In 2022 draft (see also Appendix D of the draft IRIS assessment) |
|---|---|
| | Table 2-1), with the methods and considerations described in Sections 2.1.1 and 2.1.2.<br>• Rationales for selecting certain health effect-specific values (osRfCs) and the RfC from the candidate values are presented in figures, tables, and text in Sections 2.1.3 and 2.1.4.<br>• Likewise, the rationale for choice of studies, outcomes, and datasets for use in deriving the IUR is described using text and tables in Sections 2.2.1 (nasal cancer) and 2.2.2 (myeloid leukemia). |
| Sixth, the weight-of-evidence descriptions need to indicate the various determinants of "weight." The reader needs to be able to understand what elements (such as consistency) were emphasized in synthesizing the evidence." | • The various determinants of "weight" and approaches for evaluating each determinant are described in the assessment Preface (see Table III and associated text).<br>• The way in which the evaluation of these determinants informs the evidence judgments is outlined in Preface Tables IV, V, VI, VII, VIII, and IX.<br>• In each health effect-specific section, the synthesis describes the analyses of the various determinants by discipline and the integration narrative summarizes which determinants increased or decreased certainty in the evidence overall and why, with a summary depiction of these decisions in evidence integration summary tables (e.g., Table 1-30). |

[a]From page 14 of the NAS 2011 report (these six critical recommendations are similarly described in the "Roadmap for Revision" outlined in Chapter 7, beginning on p. 151).

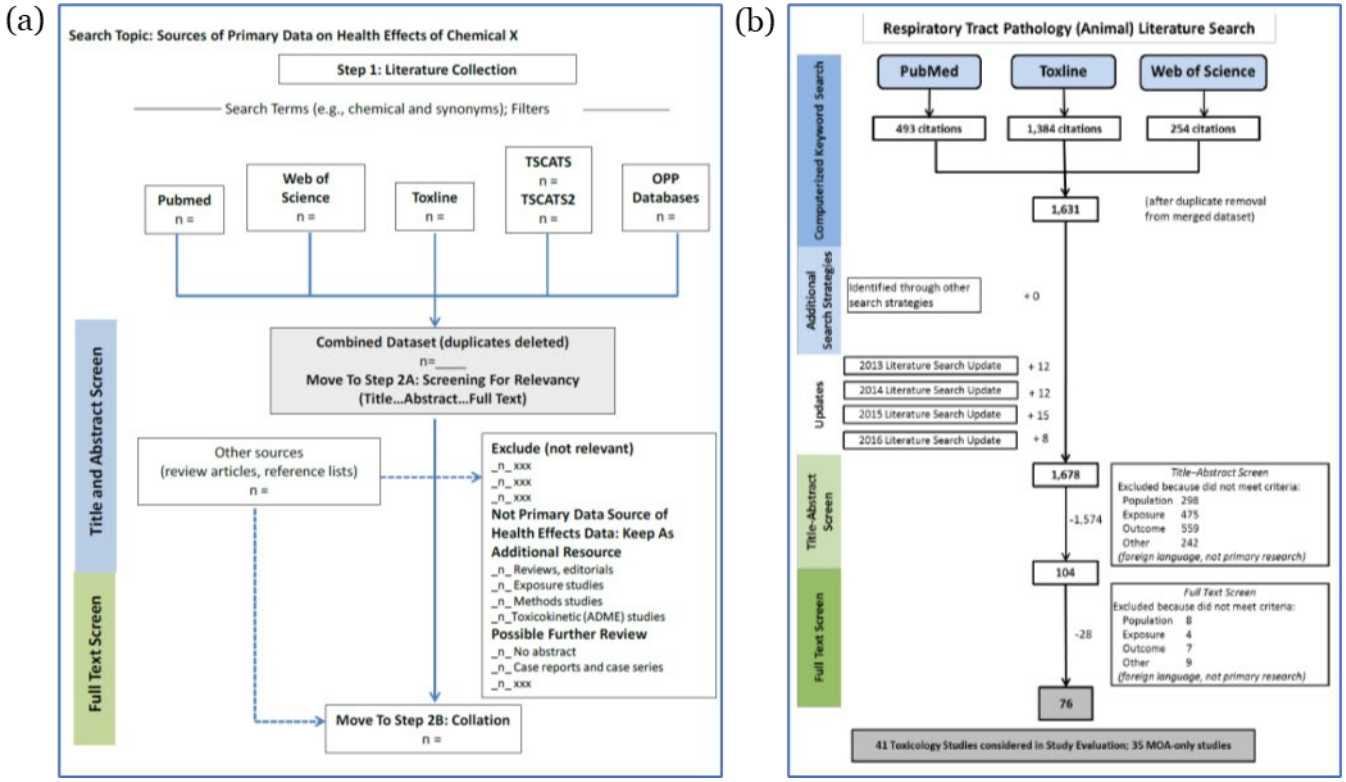


*Figure 3. Comparison of 2014 (a) and draft IRIS formaldehyde assessment (b) Literature Flow Diagrams*

*Table 5. Comparison of Study Evaluation Domains*

| **Observational Epidemiology Studies** | | | **Experimental Animal Studies** | | |
|---|---|---|---|---|---|
| Draft Assessment Domains (A.5.1) | 2014 Materials[2] (Domains) | NAS (2011) Advice[1] (Example criteria, p. 158-9) | Draft Assessment (Domains, A.5.1) | 2014 Materials[2] (Domains) | NAS (2011) Advice[1] (Example criteria, p. 158-9) |
| Participant Selection and Comparability | Participants, Selection, Follow-up | "Approach used to identify the study population and the potential for selection bias." | Exposure Quality (methods and documentation in Appendix A.5.1) | Exposure Quality | "Dosing information (dose spacing, dose duration, and route of exposure)" |
| | Comparability | "Study population characteristics and the generalizability of findings to other populations." | | | |
| Exposure measure and range | Exposure measure and range | "Approach used for exposure assessment and the potential for information bias, whether differential (nonrandom) or nondifferential (random)." | Test Subjects | Test Animals | "The species and sex of animals studied" |
| Outcome measure | Outcome measure | "Approach used for outcome identification and any potential bias." | Study Design | Study Design | "End points considered…" |
| Consideration of Likely confounding | Consideration of likely confounding | "Potential for confounding to have influenced the findings." | Endpoint Evaluation | Endpoint Evaluation | |
| Analysis and completeness of results | Analysis and presentation of results | "Appropriateness of analytic methods used." "Availability of an exposure metric that is used to model the severity of adverse | Data Considerations and Statistical Analyses | Data Presentation | |

| Observational Epidemiology Studies | | | Experimental Animal Studies | | |
|---|---|---|---|---|---|
| | | response associated with a gradient of exposures." | | | |
| Size | Sample size and Power | | [reporting issues addressed in above domains] | Reporting | |
| N/A (addressed during dose-response analysis) | - | "Precision of estimates of effect." | N/A (addressed during evidence synthesis) | - | "…, and the relevance of the end points to human end points of concern." |

[1]Materials provided to 2014 NASEM committee in Appendix F (U.S. EPA, 2013, HERO ID 1511259).

3) Beyond the updating of the human and animal evidence, was the search for mechanistic evidence updated from 2017 forward? If so, how were impactful mechanistic studies identified?

EPA Response to Question 3:

Yes, the same search strategies (and screening against PECO criteria) applied through 2017 were applied in updating the literature through 2021 using the systematic evidence mapping approach outlined in Appendix F. This included the searches for mechanistic information across the various noncancer health effect-specific search strategies, the specific searches for mechanistic studies relevant to upper respiratory tract and lymphohematopoietic cancers, as well as the augmented search for mechanistic information related to inflammation and immune effects (see Appendix Table F-2). The screening decisions are documented in HERO and through interactive HAWC literature trees on each search strategy (see links in Appendix F).

The considerations for identifying studies potentially impactful to the 2017 draft assessment conclusions are outlined in the section titled, "Considerations for identifying possibly impactful studies" in Appendix F. Broadly, 'possibly impactful studies' were those interpreted as having the potential to affect key assessment conclusions (i.e., hazard judgments or toxicity values). As mechanistic studies can include animal and human studies, as well as in vitro studies and computational models, the considerations relevant to specific study types (e.g., human or animal studies) outlined in this section were applied to any such identified mechanistic studies. In addition, the following consideration more specific to mechanistic studies was applied: "More apical endpoints and those most directly related to the mechanistic uncertainties identified in the 2017 draft as most relevant to drawing hazard or dose-response judgments were considered more impactful. The specifics of this consideration vary depending on the health outcome(s) of interest. In some cases, this relevance determination relates to the potential human relevance of the endpoints, while in others this relates to an ability to infer adversity."

For each of the individual literature search updates, there is a corresponding section of Appendix F that specifies the studies meeting the PECO criteria and indicates whether and why each study was deemed to be possibly impactful to the 2017 draft conclusions (with the possibly impactful studies being incorporated in the external review draft). See Table 6 for a characterization of the results specific to mechanistic studies, which illustrates that approximately 31 mechanistic studies published after 2017 were identified as "possibly impactful" and incorporated into the current external review draft assessment.

*Table 6. Mechanistic Studies Identified in the SEM and Their Disposition*

| Search[a] | Met PECO (possibly impactful) | Includes "mechanistic" data | Example rationales for not impactful |
|---|---|---|---|
| Mechanistic Studies of Upper Respiratory Tract Cancer[b] | 27 (8) | 27 (8) | Nonspecific biomarkers; formalin or unknown test article; in vitro (low relevance to inhalation focus) |
| Mechanistic Studies of Lymphohematopoietic Cancer[b] | 25 (14) | 25 (14) | Not key endpoint impacting conclusions; formalin test article, high levels; non-primary research |
| Inflammation and Immune Effects (mechanistic information) | 56 (8) | 56 (8) | Indirect ROS measures not key endpoints impacting conclusions;<br>formalin or unknown test article (confounding); acute duration |
| Developmental and Reproductive Toxicity | 9 (5) | 5 (1) | Formalin or unknown test article |
| Nervous System Effects | 14 (2) | 8 (0) | Formalin or unknown test article; high levels; lack of vehicle control |

[a]Several searches were specific to study types other than mechanistic studies (e.g., upper respiratory tract cancer in animals); these searches are not included in this table.

[b]The searches for cancer mechanisms primarily focused on genotoxicity endpoints; the searches for mechanistic research on inflammation and immune effects and respiratory pathology retrieved studies also relevant to cancer.

4) EPA noted that study authors were contacted for "key study details". Did EPA ask for data on all medium/high confidence studies that would enable them to derive POD? For example, Liu *et al*. 1991 was identified as medium confidence but the study was not included in the POD derivation because of "incomplete reporting of modeling results" (Table 2-1).

EPA Response to Question 4:

As described in several places in the draft IRIS formaldehyde assessment, EPA sought additional study details when such information might have improved the clarity of EPA's understanding, and not solely for POD derivation. EPA noted in the *Preface* section on *Study Evaluation*, "In some situations, in which key study details or results were not presented, the study author(s) were contacted to obtain this information. Any additional study details obtained from the authors are noted in the evaluation summary tables and evidence tables." In response to this question, EPA has documented 22 instances when EPA communicated with study authors by email; in 18 instances additional information was received by EPA and in 4 instances EPA noted that the authors did not respond (see Table 7).

However, EPA did not ask for data on all *medium* and *high* confidence studies that would enable the derivations of PODs. As described in the Preface and Section 2.1.1, in addition to the hazard judgments and study confidence ratings, several other considerations were applied in selecting studies for use in deriving PODs. As discussed in Section 2.1.1, selection of studies for POD derivation included an emphasis on "…use of *high* or *medium* confidence studies with appropriate study designs, complete reporting of results, and results that would not be reasonably explained by selection bias or information bias or altered by adjustment for confounding." This section goes on to indicate a preference for examination across multiple exposure levels and analyses of data at lower exposure levels. In this way, particularly when many studies were available on the same outcome, preferred studies within a given health domain were advanced over other studies supportive of those preferred studies.

There are three instances in the draft Toxicological Review where "incomplete reporting of modeling results" was indicated as the reason studies were not used to derive PODs. All are on page 2-4. The studies are Liu et al. (1991) on sensory irritation, and two pulmonary function studies by Malaka and Kodama (1990) and Wallner et al. (2012).

Liu et al. (1991) was a residential study on sensory irritation, fairly comparable in design to the primary study on this outcome that was advanced for POD derivation, Hanrahan et al. (1984). Both were *medium* confidence studies and had comparable ranges of air formaldehyde concentrations (see Figure 1-3 and discussion on p. 2-8 of

the draft Toxicological Review). Although a POD from Liu et al. (1991) was ultimately not derived due to the identified reporting deficiencies, sufficient information was available to allow for a comparison of the dose-response data to serve as a check on the POD derived from Hanrahan et al. (1984), and the Liu et al. (1991) data were found to be supportive. Given the apparent similar, and no more precise, POD that would be derived for Liu et al. (1991) if the modeling details were provided, as well as the difficulty associated with acquiring such information (note: contact information in this 20+ year-old study did not include an email address), this may explain why additional details were not sought.

For the studies on pulmonary function, confidence in the primary study advanced for POD derivation on this endpoint by Krzyzanowski et al. (1990) was *high*, as was the estimated POD for pulmonary function from Krzyzanowski et al. (1990). The studies by Malaka and Kodoma (1990) and Wallner et al. (2012) were both *medium* confidence. The mean concentration of formaldehyde in Krzyzanowski et al. (1990) was 32 ug/$m^3$ with 84% of samples below 40 ppb. Exposures were much higher at an average of 1.41 mg/$m^3$ in Malaka and Kodoma (1990) and were similar for Wallner et al. (2012) which had a median concentration of 29.8 ug/$m^3$. Based on the considerations described above, Malaka and Kodoma (1990) is clearly inferior to Krzyzanowski et al. (1990) for dose-response analysis and thus would not be preferred for POD derivation even if additional study details were made available (note: contact information in this 20+ year-old study did not include an email address). Had the missing data been made available from Wallner et al. (2012), it may have been advanced for POD derivation. However, Wallner et al. (2012) was of lower confidence due to concern for potential confounding and would be unlikely to yield a POD with equal confidence to that from Krzyzanowski et al. (1990), and thus the selected POD for pulmonary function would be unchanged. This may explain why additional details were not sought.

As illustrated in Table 7, study details were requested primarily to facilitate quantitative estimates or selection of N/LOAEL values (e.g., human and animal cancer data; human allergic responses and asthma data), although there were a few instances of requests for methodological details critical to study evaluation interpretations (e.g., applying a rating of *high* or *medium* rather than *low* confidence), including when there were only a few well-conducted studies on an outcome (e.g., nervous system effects in rodents), or when the topic was one of known or expected scientific disagreement (e.g., formaldehyde in exhaled breath; ascertainment of asthma in epidemiology studies). This approach is consistent with the description in the posted IRIS Handbook.

*Table 7. Examples of Details of EPA requests for additional information*

| Study | Author response | Documentation (page#) | Context |
|---|---|---|---|
| Aslan et al. (2006) | Yes | 1-92 | Sex and cohort information to inform whether litter effects might be influential or adjusted for in the only M or H confidence studies on an effect |
| Beane Freeman et al. (2009) | Yes | 2-83 | P-values not included in the publication |
| Beane Freeman et al. (2009) | Yes | 2-85; 2-86 | Regression parameters not included in the publication |
| Beane Freeman et al. (2009) | Yes | D-37 | Evidence of non-linearity |
| Beane Freeman et al. (2013) | Yes | 2-49 | Regression parameters (and their standard errors) from the trend tests for NPC and the cumulative exposure metric for all person-years and for exposed person-years only |
| Cap et al. (2008) | Yes | A-57 | Smoker data and ambient formaldehyde concentrations |
| Choi et al. (2009) | No | A-346 | Sampling time |
| Conolly (2003, 2004) | Yes | B-40 | Animal tumor data and dosimetry modeling |
| Dannemiller et al. (2013) | Yes | A-355 | Details on the consideration of likely confounding |
| Franklin et al. (2000) | Yes | A-309 | Details on the exposures |
| Hauptmann et al. (2004) | Yes | D-37 | Evidence of non-linearity |
| Matsunaga et al. (2012) | Yes | 1-92 | Midpoint of an exposure category, potentially to inform N/LOAEL selection |
| Neghab et al. (2011) | Yes | A-325 | Standard error of regression coefficients |
| NTP (2005) | Yes | B-48 | Animal tumor incidence |

| Study | Author response | Documentation (page#) | Context |
|---|---|---|---|
| Sarsilmaz et al. (2007) | Yes | A-620 | Sex and cohort information to inform whether litter effects might be influential or adjusted for in the only M or H confidence studies on an effect |
| Smedje and Norback (2001) | Yes | A-377 | Details of the exposure distribution and how values beneath the limit of detection were treated. |
| Subramanian et al. (2007) | Yes | B-40 | Details of the timing of expect death following tumor observation |
| Tavernier et al. (2006) | No | 1-106; A-346; A-378 | Details of the distribution of exposure levels |
| Venn et al. (2003) | Yes | 1-101; 1-113; 2-14; B-16 | To identify median or midpoint values in an exposure category for modeling or N/LOAEL selection |
| Zhai et al. (2013) | No | A-346; A-381 | Details on sampling time |
| Zhao et al. (2008) | No | A-338 | Details on exposure levels |

## EPA Addendum

This addendum provides additional details relevant to the EPA responses, most notably to Questions 1 and 2. This includes details on the information presented in EPA Figure 2 (see Table A.1), public meetings related to the draft IRIS formaldehyde assessment and evolving systematic review approaches within the IRIS Program (Table A.2), and publications related to systematic review coauthored by team members on the draft IRIS formaldehyde assessment around the time the draft was being developed before its suspension (Table A.3).

*Table A.1. Additional Details and Links on EPA Figure 2 (Timeline)*

| | |
|---|---|
| Preamble | Preamble first released in ammonia and TMBs drafts (June 2012), Link to TMBs: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=506876<br>Also, Appendix B in materials to NASEM for the 2014 review {U.S. EPA, 2013, 1511259} |
| | "Preamble to IRIS Toxicological Reviews" included: 1) Scope of IRIS; 2) IRIS process; 3) Identifying and selecting pertinent studies; 4) Evaluating the quality (and reporting the results) of individual studies (listing factors considered); 5) Weighing the overall evidence of each effect (describing the Hill considerations); 6) Selecting studies for derivation of toxicity values; and 7) Deriving toxicity values. |
| B[a]P | Public draft of B[a]P released in 2014: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=520493 |
| | Included preamble and front matter methods on "literature search" and "study selection", but did not include criteria or specific considerations for the latter. No protocol released. |
| Workshop | Formaldehyde Workshop in 2014:<br>https://www.epa.gov/iris/formaldehyde-workshop |
| | Topics: 1. Epidemiological research examining-lymphohematopoietic cancers (leukemias and lymphomas); 2. Mechanistic evidence relevant to these types of cancers; and 3.The influence of formaldehyde that is produced endogenously. |
| RDX | Public comment draft on RDX in 2016: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=527282 |
| | Included Preamble, but also included more front matter specifics on literature searches and "study selection and evaluation" (e.g., domains for animal studies were test animal; experimental design; exposure; endpoint evaluation; and results presentation, essentially the same as used in the draft formaldehyde assessment but without specific confidence ratings or documentation). Did not address evidence synthesis or integration frameworks. No protocol released. |
| ETBE and TBA | June 2017 release of ETBE and TBA; Link to TBA: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=531515 |
| | Included Preamble and additional front matter on literature search and "study selection and evaluation" similar to RDX. No protocols released. |

| | |
|---|---|
| First IAPs | First IAPs released in Sept 2017: https://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=532695 |
| First SR Protocol | First IRIS protocol release in early 2018: https://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=534484 |

*Table A.2. Public Meetings on Formaldehyde or Reviews of IRIS and its Evolving SR Approaches*

| Date | Name of Workshop/Meeting | Chemicals/Topic |
|---|---|---|
| 12/22/2022 | NAS Review of 2022 Draft Formaldehyde Assessment | NAS Review (Formaldehyde) |
| 10/12/2022 | NAS Review of 2022 Draft Formaldehyde Assessment | NAS Review (Formaldehyde) |
| 5/25/2022 | NAS Artificial Intelligence and Open Data Practices in Chemical Hazard Assessment Workshop | NAS Workshop |
| 5/9/2022 | NAS Workshops to Support EPA's Development of Human Health Assessments: Triangulation of Evidence in Environmental Epidemiology | NAS Workshop |
| 4/16/2021 | Review of the draft IRIS Handbook (Second Session) | IRIS Handbook |
| 2/11/2021 | Review of the draft IRIS Handbook (First Session) | IRIS Handbook |
| 6/3-6/4/2019 | NAS to Discuss Evidence Integration in Systematic Review: Workshop | Systematic Review |
| 12/10-12/11/2018 | NAS Strategies and Tools for Conducting Systematic Reviews of Mechanistic Data | Systematic Review |
| 2/1-2/2/2018 | NAS Review of Advances Made to the IRIS Process: A Workshop | NAS Review (IRIS) |
| 8/27/2017 | SAB Meeting on Evolution of IRIS Assessment Practices | EPA Meeting |
| 8/15-8/17/2017 | SAB External Peer Review Meeting on ETBE and TBA | SAB Review (ETBE and TBA*) |
| 12/12-12/14/2016 | SAB External Peer Review Meeting on RDX | SAB Review (RDX*) |
| 1/27-1/29/2016 | Temporal Exposure Issues Workshop | EPA Workshop |
| 12/16-12/17/2015 | Advancing Systematic Review Workshop | Systematic Review |
| 12/10-12/11/2015 | Model Averaging Workshop | EPA Workshop |
| 9/2-9/3/2015 | Epigenetics and Cumulative Risk Assessment Workshop | EPA Workshop |

| Date | Name of Workshop/Meeting | Chemicals/Topic |
|---|---|---|
| 10/15-10/16/2014 | IRIS NRC Recommendations Workshop | NAS Review (IRIS) |
| 9/2/2015 | SAB External Peer Review Meeting for Benzo[a]pyrene | Benzo[a]pyrene* |
| 7/14-7/16/2014 | SAB External Peer Review Meeting for Ammonia | SAB Review (Ammonia*) |
| 6/17-6/19/2014 | SAB External Peer Review Meeting for TMBs | SAB Review (TMBs*) |
| 4/30-5/1/2014 | Formaldehyde Workshop on Key Science Issues | EPA Workshop (Formaldehyde) |
| 8/26/2013 | Systematic Review Workshop | Systematic Review |
| 12/13/2012 | NAS Review of IRIS Process Meeting | NAS Review |
| 11/13/2012 | IRIS Public Meeting on Improvements to IRIS | EPA Meeting |

*Table A.3. Select Publications on Systematic Review Coauthored by Formaldehyde Team Members Around the Time of Draft Development*

| |
|---|
| Johns, L., G. Cooper, A. Galizia, L. Johns, and J. Meeker. Exposure Assessment Issues in Epidemiology Studies of Phthalates. ENVIRONMENT INTERNATIONAL. Elsevier Science Ltd, New York, NY, USA, 85: 27-39 (2015). https://doi.org/10.1016/j.envint.2015.08.005 |
| Kopylev, L., K. Christensen, J. Brown, and G. Cooper. A Systematic Review of the Association between Pleural Plaques and Changes in Lung Function. OCCUPATIONAL AND ENVIRONMENTAL MEDICINE. BMJ / British Medical Journal Publishing Group, London, UK, 72(8): 606-614, (2015). http://dx.doi.org/10.1136/oemed-2014-102468 |
| Segal, D., S. Makris, A. Kraft, M. Gilbert, D. Bergfelt, K. Raffaele, R. Blain, K. Fedak, M. Selgrade, and K. Crofton. Evaluation of the ToxRTool's ability to rate the reliability of toxicological data for human health hazard assessments [Regulatory Toxicology and Pharmacology 2015]. REGULATORY TOXICOLOGY AND PHARMACOLOGY. Elsevier Science Ltd, New York, NY, USA, 72(1): 94-101, (2015). https://doi.org/10.1016/j.yrtph.2015.03.005 |
| Smith, M., C. Gibbons, J. Fritz, D. DeMarini, J. Caldwell, R. Kavlock, and V. Cogliano. Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis. ENVIRONMENTAL HEALTH PERSPECTIVES. National Institute of Environmental Health Sciences (NIEHS), Research Triangle Park, NC, USA, 124(6): 713-721, (2016). https://doi.org/10.1289/ehp.1509912 |
| Makris, S., C. Scott, J. Fox, T. Knudsen, A. Hotchkiss, X. Arzuaga, S. Euling, C. Parsons, J. Jinot, K. Hogan, B. Abbott, S. Hunter, and M. Narotsky. A systematic evaluation of the potential effects of trichloroethylene exposure on cardiac development. REPRODUCTIVE TOXICOLOGY. Elsevier Science Ltd, New York, NY, USA, 65: 321–358, (2016). https://doi.org/10.1016/j.reprotox.2016.08.014 |

| |
|---|
| Cooper, G., R. Lunn, M. Agerstrand, B. Glenn, A. Kraft, A. Luke, J. Ratcliffe. Study sensitivity: evaluating the ability to detect effects in systematic review of chemical exposures. ENVIRONMENT INTERNATIONAL. Elsevier Science Ltd, New York, NY, USA, 92-93: 605-610, (2016). https://doi.org/10.1016/j.envint.2016.03.017 |
| Rooney, A., G. Cooper, G. Jahnke, J. Lam, R. Morgan, A. Boyles, J. Ratcliffe, A. Kraft, H. Schunemann, P. Schwingl, T. Walker, K. Thayer, and R. Lunn. How Credible are the Study Results? Evaluating and Applying Internal Validity Tools to Literature-Based Assessments of Environmental Health Hazards. ENVIRONMENT INTERNATIONAL. Elsevier B.V., Amsterdam, NETHERLANDS, 92-93: 617-29, (2016). https://doi.org/10.1016/j.envint.2016.01.005 |
| Kraft, A., and A. Davis. Quantitative meta-analytic approaches for the systematic synthesis of data and hazard identification: A case study of decreased pain sensitivity due to trimethylbenzene exposure. ENVIRONMENTAL RESEARCH. Elsevier B.V., Amsterdam, NETHERLANDS, 158: 598-609, (2017). https://doi.org/10.1016/j.envres.2017.07.017 |
| Radke-Farabaugh, E., J. Braun, J.D. Meeker, and G. Cooper. Phthalate exposure and male reproductive outcomes: A systematic review of human epidemiological evidence. ENVIRONMENT INTERNATIONAL. Elsevier B.V., Amsterdam, NETHERLANDS, 121(Part 1): 764-793, (2018). https://doi.org/10.1016/j.envint.2018.07.029 |
| Whalan, J., J. Stanek, G. Woodall, P. Reinhart, A. Galizia, B. Glenn, A. Kraft, S. Makris, and A. Jarabek. The Evaluation of Inhalation Studies for Exposure Quality: A Case Study with Formaldehyde. TOXICOLOGY LETTERS. Elsevier Science Ltd, New York, NY, USA, 312: 167-172, (2018). https://doi.org/10.1016/j.toxlet.2019.05.011 |