Summary of Tiered Recommendations

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.

Tier 1 Recommendations (1)

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.

Tier 2 Recommendations (41)

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 crosswalking 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.)

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

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.

Recommendation 2.5 (Tier 2):

- The assessment should be edited to more sharply demarcate the synthesis and 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.

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."

Recommendation 2.7 (Tier 2): EPA should consider using information from studies that are complementary to each other to derive benchmark concentrations for outcomes of interest (see also Appendix D). For example, multiple studies can be complementary by widening the exposure scale, broadening the age groups, and including vulnerable or susceptible groups.

Recommendation 2.8 (Tier 2): Given that EPA has requested additional information from some study authors, the authors of the Liu et al. (1991) study could be approached for additional information that would help EPA reconstruct an overall dose-response graph (see also Appendix D).

Recommendation 3.1 (Tier 2): To enhance transparency, the summary tables (e.g., Tables 24, 43, and 44) should explicitly identify the models used to derive flux values. Table 44 should clearly indicate whether the BBDR models used here are equivalent to Models 1 and 2 identified in Table 43 and the text. Table 46 should indicate which flux model was used.

Recommendation 3.2 (Tier 2): To increase transparency, EPA should provide additional clarification regarding its decision not to use the BBDR model to extrapolate rat nasal carcinogenicity to humans. Criteria used by EPA to determine whether the models would be adequately robust for this purpose are not readily available in the 2022 Draft Assessment. Likewise, EPA should provide additional support for its decision not to use the BBDR model (Connolly et al., 2004) for this extrapolation because of its conclusion that formaldehyde-induced mutagenicity, modeled as proportional to DPC concentration, is not relevant to formaldehyde's carcinogenicity.

Recommendation 3.3 (Tier 2): To increase transparency, EPA should address these shortcomings by updating tables and text to better document its dosimetry methods.

Recommendation 4.1 (Tier 2): To maximize transparency and facilitate replication, EPA should clarify the Medical Subject Headings (MeSH) terms used, list and justify any MeSH

terms that were excluded (e.g., eye, ear, nose, or skin), provide the list of national and international reviews and assessments used to identify additional references, and provide more specific links to the Health & Environmental Research Online (HERO) database where the screening decisions are documented (see Appendix A, Section A-232, line 4).

Recommendation 4.2 (Tier 2): EPA should include the body of evidence from outdoor exposure studies at the preliminary stage to derive a more holistic and objective assessment of the scientific literature.

Recommendation 4.3 (Tier 2): EPA should explicitly state what constitutes an adequate assessment of outcomes when a questionnaire is not cited, and explicitly provide the criteria used to determine the adequacy of a questionnaire. Information on these aspects of outcome assessment would facilitate replication of the EPA approach. It would be preferable for EPA to use age categories generally instead of ambiguous descriptors.

Recommendation 4.4 (Tier 2): To increase transparency, EPA should document how it assessed the potential for different types of biases, the directionality of resulting biases, and the number of biases, and state how each combination should be interpreted in terms of *high, medium, low*, or *not informative* study confidence.

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.

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 4.8 (Tier 2): If the Assessment Overview is retained (see Recommendations 2.1 and 2.2 in Chapter 2), EPA should harmonize its presentation of evidence synthesis with the presentation in the Main Assessment. In particular, the evidence synthesis section of the Assessment Overview could be updated to build upon the first three paragraphs of the "Integrated Summary of Evidence for Pulmonary Function" section in the Main Assessment (PDF p. 134).

Recommendation 4.9 (Tier 2): EPA should provide additional justification for why the most vulnerable subpopulations were not used for risk estimation, and should consider using the data from children with asthma that are provided in Krzyzanowski et al. (1990).

Recommendation 4.10 (Tier 2): EPA should provide an explicit description of the comparator used in screening human and animal studies, and resolve discrepancies between search terms and inclusion and exclusion criteria.

Recommendation 4.11 (Tier 2): EPA should provide a consistent rationale for each study quality domain used in the assessment.

Recommendation 4.12 (Tier 2): EPA should provide a comprehensive description or listing of immunopathologies that were considered as potentially related to formaldehyde before the decision was made to limit the focus to prevalent allergies and prevalent asthma.

Recommendation 4.13 (Tier 2): EPA should explicitly state its rationale for age based exclusions and define the terms "very young" and "very old," better justify the rationale for excluding allergic contact dermatitis and food allergy as outcomes of interest, and provide the rationale for excluding animal models of "the development of immunological or allergy" outcomes (unless such studies do not exist or are inadequate).

Recommendation 4.14 (Tier 2): EPA should include a specific statement on the age at which asthma diagnosis is considered valid to justify the age exclusions for young children, as well as the category of "the elderly".

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).

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.).

Recommendation 5.2 (Tier 2): For consistency, EPA should add a summary of the data and evidence synthesis for laryngeal cancer after page 103 of the Assessment Overview.

Recommendation 5.3 (Tier 2): To add clarity, EPA should, in the captions of figures displaying the findings of epidemiological studies for the different cancers, provide references to the

numbers of the tables that describe the confidence in each study (*low*, *medium*, or *high*) and "results" (high vs. medium confidence, as presented in Figure 1-38).

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.

Recommendation 5.5 (Tier 2): EPA should acknowledge the uncertainty involved in interpreting the analyses on the degree to which exposure-response relationships are stronger than cumulative exposure for determination of peak exposure and risk.

Recommendation 5.6 (Tier 2): EPA should state how the adjustment for the 15-year lag was made for nasopharyngeal cancer mortality, and explain the assignment of zero exposure in Table B-18.

Recommendation 5.7 (Tier 2): In Appendix B, Table B-12, increasing the number of significant figures in columns P, H, and L to align with column I would add transparency by making it easier for readers to follow and understand the calculations for nasopharyngeal cancer incidence.

Recommendation 5.8 (Tier 2): EPA should describe more clearly the procedure and justification for pooling the data from two animal studies into one analysis, and clarify that combined and corrected incidence data are contained in the Bermudez memorandum, which is not readily accessible to the public. The individual animal data for time-to-tumor occurrence used in the model should be provided in an appendix.

Recommendation 5.9 (Tier 2): To enhance transparency, EPA should provide additional detail on the modeling, including constraints imposed on model parameters, the results of model fitting (goodness-of-fit test), and the approach used to define lag parameters. The relationship between administered dose and the DNA– protein crosslinks and flux dose metrics should also be provided. Given the uncertainties in the dose surrogates, a dose-response analysis and benchmark concentration calculations using administered concentrations should be provided as a point of comparison.

Recommendation 5.10 (Tier 2): EPA should organize the discussion of uncertainties and variabilities in a manner that is easier to follow, such as by models or by process (models, benchmark concentration estimation, lower dose extrapolation, or extrapolation from animal data to humans).

Recommendation 5.11 (Tier 2): The results from different models and different databases are remarkably similar, supporting each other and suggesting a good degree of robustness. EPA should highlight this robustness to a greater degree while not losing sight of uncertainties within individual studies, endpoints, and models.

Recommendation 5.12 (Tier 2): EPA should discuss the extent to which the inhalation unit risk estimates based on animal squamous cell carcinoma data and mechanistic data provide supporting evidence for the inhalation unit risk based on the human nasopharyngeal carcinoma data.

Recommendation 5.13 (Tier 2): EPA should address technical errors, such as mischaracterization of a trend p-value, with a thorough and technical edit and proofreading.

Recommendation 5.14 (Tier 2): EPA should discuss the implications and interpretation of nonmonotonic dose-response relationships observed with the cumulative exposure metric (e.g., p. 2-92, lines 2–4).

Recommendation 5.15 (Tier 2): In the discussion of uncertainties and confidence in the inhalation unit risk for myeloid leukemia, EPA should include the unknown dose rate-response relationship, the choice of statistical model and method, and the lack of understanding of mechanism. The three estimates in Table 2-35 should be presented as alternative, low-confidence inhalation unit risk estimates for myeloid leukemia without selection of a preferred estimate. EPA should not characterize the combining of other/unspecified leukemia with myeloid leukemia as "the best approach."

Recommendation 5.17 (Tier 2): For clarity, EPA should include the lifetable calculations for the adult-only unit risk estimate in Appendix B.

Recommendation (Tier 2): EPA should clearly articulate what is meant by "appropriate exposure circumstances" in Section 2 or abandon the use of the term.

Tier 3 Recommendations (4)

Recommendation 4.5 (Tier 3): In the population selection criteria, the potential for selection bias could be assessed by considering the proportion of the eligible population invited to participate in the study and the proportion of the eligible population that was ultimately included in the analysis. EPA should state the criteria used to assess selection bias in the text, tables, and figures.

Recommendation 4.15 (Tier 3): EPA should include studies with headache as an outcome to maintain consistency with other health effect categories. Alternatively, a stronger rationale should be provided for exclusion of headache other than its perceived subjectivity. Headache could be combined with other self-reported neurotoxicity outcomes.

(follows recommendations 4.16 (tier 2))
Additionally, EPA should address the following points (Tier 3):

- Handle dose-response modeling of correlated data (e.g., Andersen and Molhave, 1983; Kulle et al., 1987) by standard statistical methods, employing a two-step process that involves first fitting a dose-response model for correlated data using standard statistical methods, and then deriving BMC and BMCL using the fitted model.

- Develop methodology that goes beyond a qualitative display of the variability and uncertainty of cRfCs or osRfCs. The current EPA method has limited reproducibility because of the lack of detail. A meta-analysis approach offers a viable option.

Recommendation 5.16 (Tier 3): Terminology for inhalation unit risk estimates and for potency values used in applying ADAFs should be consistent across the IRIS Program, including with terms in the IRIS glossary.