EPA 600R26001 | July 2026

PREAMBLE TO THE INTEGRATED SCIENCE ASSESSMENTS

U.S. Environmental Protection Agency
Office of Clean Air Programs
Impacts and Ambient Standards Division
Research Triangle Park, NC

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

The Environmental Protection Agency's (EPA) Integrated Science Assessments (ISAs) are intended to concisely review, synthesize, and evaluate relevant scientific information and reach key science judgments to inform the Agency's periodic reviews of the National Ambient Air Quality Standards (NAAQS) as required by the Clean Air Act (CAA). Section 108 of the CAA requires that "[a]ir quality criteria for an air pollutant shall accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of [the] pollutant in the ambient air…". The ISAs document air quality criteria and form the scientific foundation for establishing and reviewing primary and secondary NAAQS to protect public health and public welfare, respectively. The ISAs, in conjunction with additional technical analyses and policy assessments, inform proposed and final decisions on the NAAQS for the six criteria air pollutants: ozone (O_3) and related photochemical oxidants, oxides of nitrogen, sulfur oxides, carbon monoxide (CO), lead (Pb), and particulate matter (PM).

This document describes the EPA's approach to developing ISAs, replacing the 2015 Preamble to the ISAs (U.S. EPA, 2015). Updates since the 2015 Preamble reflect advancements presented in recently completed ISAs (e.g., U.S. EPA, 2019a; U.S. EPA, 2020b; U.S. EPA, 2024a); recommendations made by an ad hoc committee of the National Academies of Science, Engineering, and Medicine (NASEM) charged with reviewing the ISA framework for causality determinations (NASEM, 2022); and the evidence assessment approaches described in recent ISA planning documents (U.S. EPA, 2024b, Appendix A; U.S. EPA, 2024c, Appendix A).

The key steps in ISA development are summarized in Figure 1 and described in detail below in sections 2 through 6.[1] These sections describe the approaches to defining the ISA scope (section 2), conducting the literature search and initial study screening (section 3), evaluating individual study quality (section 4), synthesizing and

[1] Consistent with NASEM recommendations (e.g., NASEM, 2022, p. 132), the approach described in this updated Preamble may evolve with advancements in evidence integration or weight-of-evidence evaluations.

integrating the evidence to inform ISA conclusions (section 5), and obtaining scientific and public review of the draft ISA (section 6).

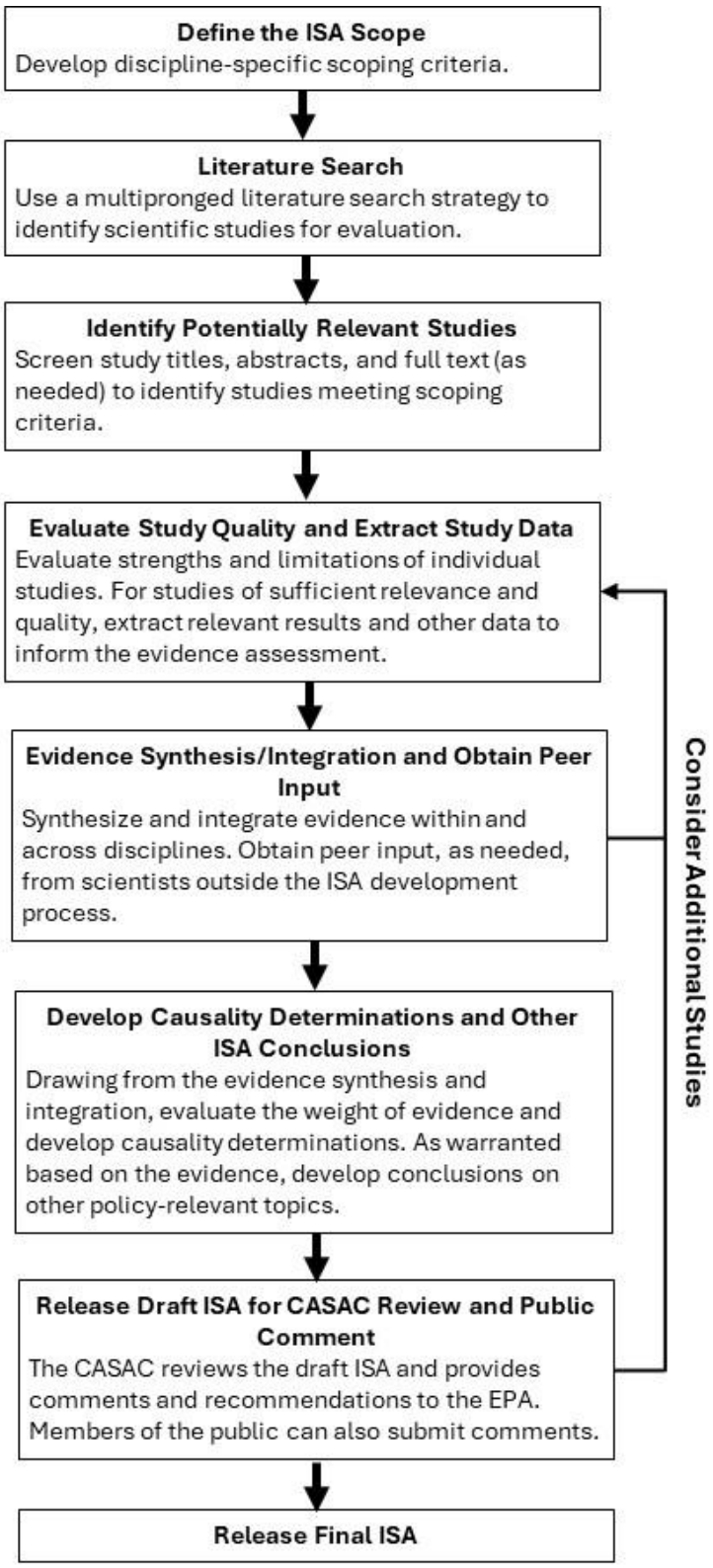

Figure 1. Overview of Integrated Science Assessment Development.

2. DEFINING THE SCOPE OF THE ISA

Integrated Science Assessments present an evaluation of the scientific information relevant to the subject air pollutant and the effects of its presence in ambient air. This evaluation is documented in a series of topic-specific sections that integrate recent studies (i.e., those that have become available since the previous ISA) with older studies, building on the information presented in previous assessments. The ISAs evaluate and describe the degree to which recent studies address scientific uncertainties and limitations identified in prior NAAQS reviews. Important older studies (e.g., those included in previous ISAs) may also be discussed in detail to reinforce key concepts and conclusions, and these older studies may be reinterpreted in the context of more recent data. Older studies may be the primary focus of the ISA in some subject areas or scientific disciplines where research efforts have subsided or where those older studies remain the definitive works available in the literature. The ISAs emphasize studies that are most relevant for ambient air-related exposures in the United States, including studies examining pollutant concentrations that reflect the range of those exposures across microenvironments. Studies examining higher exposure concentrations may be included if they evaluate exposure-response relationships across a range of exposures that also include levels relevant to ambient air concentrations, if they utilize novel or particularly strong study designs, or if they provide evidence of the potential biological mechanism(s) for an observed effect.

To be included in an ISA, relevant studies must have undergone scientific peer review and generally should have been published or accepted for publication within the predefined literature search cutoff dates.[2] The ISA uses discipline-specific scoping statements to identify potentially relevant studies. These scoping statements (presented in the following sections) define the objectives of the evidence assessment and establish criteria that should be met to consider a study for inclusion in the ISA. Included studies must present original research or new analyses of existing data. Review articles that are out of scope or that are limited to qualitatively summarizing and interpreting existing

[2] Studies that fall outside stated literature cutoff dates can be considered for inclusion in the ISA if recommended by the CASAC, submitted in public comments, or identified as a result of peer input.

studies, without presenting new information or analyses, are not considered for inclusion in ISAs. General criteria that guide the ISA scoping decisions for studies of atmospheric and other environmental sciences, exposure science, human health effects, and welfare effects are discussed below in sections 2.1 to 2.4.

2.1. Atmospheric and Other Environmental Science Studies

Atmospheric science studies are included in ISAs to provide updated information on the sources and environmental factors that influence surface concentrations of criteria pollutants in ambient air. Some criteria air pollutants (e.g., Pb, sulfur oxides, oxides of nitrogen, PM) can pass through media other than air (e.g., deposited in water, soil), and for these pollutants environmental science studies other than atmospheric science may also be included in an ISA. Information from atmospheric science studies, and from other environmental science studies when appropriate, can provide important context for interpreting the health and welfare effects evidence, particularly when those studies provide insight into the spatiotemporal variation in ambient concentrations.[3] Relevant studies are identified using scoping statements that define study inclusion criteria in terms of pollutant sources, transport and transformation, extent, and measurement and modeling (STEM) (Table 1). Studies meeting inclusion criteria for any of the four aspects of the STEM statement are considered for inclusion in the ISA. Application of STEM statements is consistent with ISA scoping statements used for the health and welfare effects evidence (see Sections 2.2 to 2.4) and with NASEM recommendations to provide clarity in the study selection and evaluation process (NASEM, 2022).

For atmospheric and other environmental sciences, flexibility is built into the application of STEM statements by maintaining a broad scope for subject areas with few studies identified and dynamically narrowing the scope (e.g., in areas with more studies) to account for novel scientific findings, geographic similarity to the United States, representativeness or diversity of environmental conditions, quality of measurement or

[3] In the ISA atmospheric science sections, results of peer-reviewed published studies are often supplemented by additional relevant EPA analyses and data products, for example to illustrate trends in precursor emissions or ambient air concentrations.

modeling method used, or other refinements. The STEM statements for a specific ISA are informed by the evidence from previous ISAs and Air Quality Criteria Documents (AQCDs) and by expert knowledge of the relevant scientific literature. These scoping statements can serve to highlight well-established areas of research, as well as gaps in the literature and uncertainties from previous assessments (U.S EPA, 2020a; U.S. EPA, 2016). A generic STEM statement for atmospheric science studies is provided below (Table 1). This generic statement provides a scoping framework that can be modified as appropriate for specific ISAs.[4]

Table 1. Generic STEM statement for identifying relevant atmospheric science studies.

Statement	Description
Source (S)	Emissions estimates of the criteria pollutant and/or precursors to the criteria pollutant, as well as observations or modeling of physical and chemical characteristics relevant to understanding sources of the criteria pollutant and/or precursors to the criteria pollutant in the atmosphere.
Transport and Transformation (T)	Chemical transformations that lead to the production and destruction of the criteria pollutant in the atmosphere, as well as atmospheric transport processes (at various spatial scales and including deposition) that influence ambient concentrations of the criteria pollutant.
Extent (E)	Observations and modeling estimates of ambient air concentrations and trends of the criteria pollutant in the United States, including its spatial variability over various scales (e.g., global, national, regional) and temporal trends (e.g., diurnal, seasonal, long-term); or characteristics, such as composition or relationship with atmospheric properties that are relevant to understanding spatiotemporal concentration trends of the criteria pollutant.
Measurement and Modeling (M)	Studies describing methods to measure or model ambient concentrations of the criteria pollutant or precursors/products of the criteria pollutant, including the evaluation of measurement principles and modeling assumptions, examination of potential bias and uncertainties, and method intercomparisons that are relevant to the NAAQS or other studies in the ISA.

2.2. Exposure Science Studies

Studies on exposure science are included in the ISA to provide information on the relationship between concentrations of criteria pollutants in ambient air and human exposure to the pollutant. This information can provide important context for

[4] For criteria air pollutants that can pass through media other than air, the STEM statement(s) would identify atmospheric science studies as well as other environmental science studies, as appropriate for the pollutant.

interpreting the health effects evidence, particularly when those studies provide insight into the spatiotemporal variation in exposures, the variability and uncertainty associated with exposure estimates, or the strengths and limitations of exposure surrogates commonly used in epidemiologic studies. Potentially relevant exposure science studies are identified using STEM scoping statements that define study inclusion criteria in terms of pollutant sources, transport and transformation, exposure, and measurement and modeling (STEM) (Table 2). Studies meeting criteria for any of the four aspects of the STEM statement are considered for inclusion in the ISA. A generic STEM statement for exposure science studies is provided below (Table 2). This generic statement provides a scoping framework that can be modified as appropriate for specific ISAs.

Table 2. Generic STEM statement for identifying relevant human exposure science studies.

Statement	Description
Source (S)	Emissions from outdoor or indoor sources (e.g., traffic or gas ranges emissions), including both anthropogenic sources (e.g., industrial emissions) and natural sources (e.g., wildfires), and pollutants directly emitted or formed through chemical reactions in the atmosphere (e.g., O_3).
Transport and Transformation (T)	Atmospheric and environmental processes, including the transport of air pollutants at various scales (i.e., national/global, regional, urban, neighborhood, middle, micro scales, and microenvironments), chemical transformations (e.g., photochemical reactions), deposition, and dynamics within microenvironments (e.g., indoor chemistry).
Exposure (E)	Exposures characterized by various surrogates (e.g., ambient air measurement, near-source measurement, microenvironmental measurement, personal exposure measurement and modeling, biomarkers of exposure) and exposure determinants (i.e., factors leading to differential exposures, such as proximity to sources and activity patterns). This includes spatiotemporal trends of various exposure surrogates and populations experiencing elevated exposures or exposure patterns identified in health studies as being linked to increased health risks (e.g., due to exposure concentration, duration, and/or frequency).
Measurement and Modeling (M)	Measurement (e.g., federal reference and equivalent methods, passive samplers, remote sensing, and biomonitoring approaches) and modeling techniques (e.g., land use regression, chemical transport, microenvironmental models, and hybrid or ensemble models) characterizing ambient air, indoor/microenvironmental air, and/or personal exposures, including the evaluation of measurement principles and modeling assumptions, examination of potential bias and uncertainties, and comparison of different techniques.

2.3. Health Effects Studies

Health effects studies included in an ISA provide insight into the public health impacts of criteria pollutant exposures and into the strengths and limitations of the evidence supporting those impacts. To identify potentially relevant health studies, ISAs use discipline-specific population, exposure, comparison, outcome, study design (PECOS) statements. The PECOS statements help define the objectives of the assessment and establish relevance criteria that should be met to consider a study for inclusion in the ISA, thereby facilitating the transparent identification of potentially relevant literature (NASEM, 2022). Studies that meet all PECOS criteria during title and abstract and full text review are considered for inclusion (see section 4).

The PECOS statements are informed by the body of evidence from previous ISAs and AQCDs, as well as by expert knowledge of the relevant scientific issues. Generic examples of PECOS statements are provided below for epidemiologic (Table 3), controlled human exposure (Table 4), and animal toxicological (Table 5) studies. In each ISA, modifications of these generic PECOS statements may be included as appropriate for specific pollutants and health outcome categories.

Table 3. Generic PECOS statement for identifying relevant epidemiologic studies.

Exposure Duration	Population, Exposure, Comparison, Outcome, Study Design (PECOS)
Short-term or long-term exposure	**Population (P):** Any human population, including populations or lifestages that might be at increased risk.
	Exposure (E): Exposure to the pollutant(s) under evaluation at concentrations relevant to ambient air in the United States.[5]
	Comparison (C): Per unit increase in pollutant exposure (e.g., in ppb or $\mu g/m^3$) or populations exposed to lower concentrations of pollutant compared to higher concentrations (e.g., categorical comparisons between different exposure metric quantiles).
	Outcome (O): Change or difference in risk (e.g., incidence/prevalence) of health outcome (e.g., respiratory effects, cardiovascular effects, nervous system effects, metabolic effects, reproductive and developmental effects, cancer, total mortality) per change or difference in exposure.
	Study Design (S): Epidemiologic studies, including but not limited to panel, case-crossover, time-series, case-control, cohort, cross-sectional, and quasi-experimental studies with appropriate timing of exposure for the health outcome of interest.[6]

Table 4. Generic PECOS statement for identifying relevant controlled human exposure studies.

Exposure Duration	Population, Exposure, Comparison, Outcome, Study Design (PECOS)
Single or repeated short-term exposures	**Population (P):** Human volunteers enrolled in controlled exposure studies, including volunteers representing populations or lifestages that might be at increased risk of pollutant-related health effects.
	Exposure (E): Controlled exposure to relevant pollutant concentrations, where exposures must be controlled by the experimenters and not simply a measure of ambient air or occupational exposure.
	Comparison (C): An appropriate control exposure (e.g., filtered air or room air) for each study participant or an appropriately matched comparison group exposed to filtered air or room air.
	Outcome (O): Outcomes of interest are those that relate to human health including, but not limited to, effects on the respiratory system, cardiovascular system, immune

[5] To focus on the exposures most relevant for air quality in the United States, an upper limit may be applied during study evaluation. The definition of "relevant" exposures will differ across pollutants.

[6] Meta-analyses of existing literature may also be considered for inclusion in an ISA to the extent that (1) the time period from which the underlying literature is drawn is relevant to the ISA time period (e.g., so that the pooled effect estimates are reflective of current understanding of the literature) and (2) the underlying literature would meet the PECOS criteria.

	system, nervous system, diabetes, cancer, or reproduction and development. Effects of interest include changes in indicators or measures of physiological function, health-relevant biomarkers, and organ structure. Effects can be directly measured in exposed study participants or in cells, tissues, or fluids isolated from study participants.
	Study Design (S): Studies that perform controlled human exposures meeting the above criteria or that analyze data from previously conducted controlled human exposures (e.g., reanalysis, meta-analysis).

Table 5. Generic PECOS statement for identifying relevant animal toxicological studies.

Exposure Duration	Population, Exposure, Comparison, Outcome, Study Design (PECOS)
Short-term or long-term exposure	**Population (P):** Laboratory nonhuman mammalian animal species (e.g., nonhuman primate, mouse, rat, guinea pig, minipig, rabbit, cat, dog) of any lifestage including models of increased susceptibility.
	Exposure (E): Controlled exposure to relevant pollutant concentrations.
	Comparison (C): An appropriate control exposure (e.g. filtered air or room air).
	Outcome (O): Outcomes of interest are those that relate to human health including, but not limited to, effects on the respiratory system, cardiovascular system, immune system, nervous system, diabetes, cancer, or reproduction and development. Effects of interest include changes in indicators or measures of physiological function, health-related biomarkers, and organ structure. Effects can be directly measured in exposed animals or in cells, tissues, or fluids isolated from animals.
	Study Design (S): Controlled exposure studies of animals *in vivo* meeting the above criteria.

2.4. Ecological and Other Welfare Effects Studies

Ecological and other welfare effects studies included in an ISA provide insight into the public welfare impacts of criteria pollutant exposures and into the strengths and limitations of the evidence supporting those impacts. Ecological effects considered in the ISAs typically include several of the topics given as examples by the Clean Air Act definition in Section 302(h) related to effects on welfare including soils, water, vegetation, animals, and wildlife. Additional topic areas that may be evaluated by the ISA include impacts on visibility, weather, climate, materials damage, economic values, personal comfort, and well-being. During ISA development and scoping, targeted literature searches may be conducted to evaluate whether any new information on known non-ecological welfare effects or additional welfare effects not considered in prior ISAs has become available.

Emphasis is placed on studies that evaluate effects of exposures relevant to ambient air concentrations, unless lower or higher than ambient air concentrations are part of a set of exposures that include ambient or near-ambient exposures in an experiment designed to study exposure-response. Study design methods for welfare effects may encompass a range of experimental approaches (e.g., controlled laboratory exposure, growth chamber, open-top chamber, mesocosm, gradient, field study). Welfare effects studies may also include those that assess the relative importance or value placed by the public on protection of different uses or services.

The ISAs use the level of biological organization, exposure, comparison, endpoint, and study Design (LECES) statements to identify potentially relevant ecological studies. The LECES statement was developed by the EPA to provide a question formulation system best suited to the nature of ecology studies, comparable to PECOS in health studies and STEM in atmospheric and exposure science studies. The LECES statements work on the same principle as PECOS, in that studies must meet all scoping criteria (i.e., L/E/C/E/S) to be considered for inclusion in an ISA. A generic LECES statement to aid in identifying relevant ecological effects studies is provided below (Table 6). In each ISA, modifications of the generic LECES statement in Table 6 may be included to focus on particular study designs (e.g., experimental, observational) and on particular ecological or other welfare endpoints.

Table 6. Generic LECES statement for identifying relevant ecological studies.

Statement	Description
Level of Biological Organization	Any species, individual organism, population, biological community, or ecosystem
Exposure	Concentrations occurring in the environment, experimental concentrations within a given range of recent concentrations, or at least one concentration in the test series in the range described above.
Comparison	Relevant control sites, treatments, or parameters.
Endpoint	Effects relevant to the specific pollutant(s) being assessed (e.g., growth, survival, reproduction, community composition, productivity, ecosystem processes, or others).
Study Design	Laboratory, greenhouse, chamber, mesocosm, fumigation, field, gradient, mechanistic modeling, meta-analysis, or other studies.

3. LITERATURE SEARCH AND SCREENING

3.1. Literature Search

The EPA uses a multipronged approach to identify scientific studies that may be relevant for inclusion in a particular ISA. This approach can employ both broad and targeted literature search strategies.

An often-used strategy for literature identification starts with querying relevant databases (e.g., PubMed and Web of Science) with Boolean strings of keywords. When this approach is applied, a set of keywords is curated for each pollutant and/or topic. Keywords can be designed to return a broad set of search results (i.e. very high recall but low precision) or they can target particular disciplines or topics. The results of a broad keyword search may be categorized by scientific discipline (e.g., epidemiology, toxicology, ecology) using automatic topic classification (e.g., DoCTER, https://www.icf-docter.com/). In this classification step, a set of relevant references for each discipline is used to train machine learning algorithms. The classification algorithm then categorizes the search results into the appropriate disciplines based on information from the seed set.

A more targeted approach to literature identification that may be employed is forward citation searching. For this approach, a set of relevant published references is identified as a seed set and databases (e.g., PubMed and Web of Science) are then queried to identify recently published literature that has cited any of the references in the seed set. The seed set comprises data-containing peer-reviewed references cited in the previous ISAs or AQCDs. Seed sets are specific to topic area or discipline (e.g., atmospheric science, epidemiology). Because each seed set is highly relevant to the topic of interest, this targeted approach to reference identification can be more precise than keyword searches.

Studies identified by keyword and forward citation searches can be combined with studies identified by public commenters or during peer input, studies recommended by the CASAC (see section 6), and studies identified by EPA scientists based on professional expertise. This large group of studies is then deduplicated and used as the complete set of scientific studies to be screened for potential relevance to the ISA, as described below.

3.2. Literature Screening

Studies identified during the literature search described above are evaluated for potential relevance using a multi-tiered literature screening approach designed to maximize efficiency while simultaneously ensuring relevant studies are identified. Studies are initially evaluated for potential relevance by comparing their titles and abstracts to the discipline-specific scoping criteria defined by STEM, PECOS, or LECES statements (section 2).[7] Because the number of criteria pollutant-related studies identified in initial literature searches can be very large (e.g., hundreds of thousands of studies across disciplines and outcomes), the EPA often uses machine-assisted literature review software tools that rank references based on user input, allowing reviewers to efficiently identify potentially relevant studies (e.g., DistillerSR [Ottawa, Ontario, Canada] Hamel et al., 2020; SWIFT-Active Screener [Sciome, RTP, NC, USA] Howard et al., 2020; Living Literature Review [U.S. EPA, RTP, NC, USA] (U.S. EPA, 2023)[8]). Studies that appear to meet the ISA scoping criteria based on the title and abstract screen, together with studies that cannot be definitively identified as out of scope, are retained for further evaluation of the full text.

Relevance to the ISA is further evaluated by comparing the full text of documents to the predefined STEM, PECOS, and LECES scoping criteria described above. Studies identified as STEM-, PECOS-, or LECES-relevant are evaluated for individual study quality as described below (section 4). Studies that are determined to meet scoping criteria and judged of sufficient quality are included in the ISA. Studies that do not meet study quality criteria are excluded.

[7] For health effect studies, EPA scientists may seek supplementary scientific information to evaluate biologically plausible modes of action between air pollutant exposures and effects. Given the range of data types that could support biological plausibility, this supplementary scientific information may be derived from studies that are out of the scope of PECOS statements (e.g., exposure concentrations higher than PECOS-defined cut-offs, alternative model systems that do not meet PECOS criteria).

[8] The Living Literature Review tool (https://heronetnext.epa.gov/reference/12232868/) is part of EPA's Health and Environmental Research Online (HERO) database (https://hero.epa.gov/).

4. EVALUATING INDIVIDUAL STUDY QUALITY

Studies that are determined to be STEM-, PECOS-, or LECES-relevant are evaluated for study quality. Individual study quality is evaluated by considering the design, methods, and documentation of each study, but not study results. While results are not considered as part of a study quality evaluation, additional examination of study design and methods may be appropriate to understand results that are inconsistent with overall trends observed in a particular body of evidence. The ISA's general approach to study quality evaluation considers the strengths and limitations of individual studies, including the possible role of chance, confounding, and other biases that may affect study interpretation and the strength of inference that can be drawn from study results. Consistent with NASEM recommendations (NASEM, 2022, p. 5), the sections below identify foundational study design attributes and analysis approaches that are considered when evaluating study quality and describe how these attributes and approaches can influence the inference drawn from individual studies.

Effective evaluation of study quality relies most fundamentally on transparency, including clearly reporting key data, assumptions, methods, formulas, input parameters, quality assurance and quality control (QA/QC) procedures, statistical models/coding, reasoning process, and limitations. Such transparency can increase confidence in individual study results (NASEM, 2022, p. 6-7). Inference is stronger for studies that clearly document the primary and any secondary aims of the study, or specific hypotheses being tested. The sufficiency of study documentation and reporting can be evaluated based on (1) whether a person with a general knowledge of the research area can understand the described approach, (2) whether the study can, in principle, be independently verified or replicated based on the reported methodology, should original data be available, and (3) whether study limitations can be characterized based on the reported assumptions and uncertainties (WHO, 2008). The ISA assessment of scientific quality for individual studies is framed by the following general questions:

- Are the study design, study groups, methods, data, and results adequately justified and clearly presented in relation to the study objectives to allow for study evaluation, including evaluation of underlying assumptions and limitations?

- Are the air quality, exposure, and/or dose metrics sufficiently representative of or pertinent to ambient air and are they adequately documented?
- Do the analytical methods used in the study provide adequate sensitivity and precision?
- Are the statistical analyses appropriate, properly performed, and appropriately interpreted?
- Are the study populations, participants, organism model systems, ecosystems, or study site(s) appropriately selected and sufficiently well-defined to allow for meaningful comparisons between study or exposure groups?
- Are the health, ecological, or other welfare effect measurements meaningful and valid?
- Are the likely confounders controlled for and modifying factors examined in the study design and/or statistical analysis?
- Was the study conducted with appropriate oversight by ethics boards or committees (e.g., documenting approval from a U.S. or foreign-equivalent Institutional Review Board (IRB) for human studies or from an Institutional Animal Care and Use Committee (IACUC) for animal studies)?

Considering these general questions can inform decisions on study inclusion and provide context for weighing and integrating the evidence and developing ISA conclusions. Additional discipline-specific study quality considerations are discussed below in sections 4.1 to 4.3. These sections discuss the study attributes considered in the ISA evaluation of individual study quality for atmospheric and exposure science studies (4.1), human health studies (4.2), and ecological and other welfare studies (4.3). These attributes can be adapted for specific ISAs, as appropriate.

4.1. Atmospheric and Human Exposure Science

The ISA study quality evaluation for both atmospheric and exposure science studies considers the applicability and utility of the study, the soundness of the study approach, the clarity and completeness of the study, and its treatment of uncertainty and variability.[9] These study attributes are adopted from the EPA exposure assessment

[9] For pollutants that pass through media other than air (e.g., Pb, sulfur oxides, oxides of nitrogen, PM), an analogous approach is used to evaluate the quality of other environmental science studies.

guidelines (U.S. EPA, 2019b)[10] and can be used collectively to inform conclusions on study quality.

4.1.1. Applicability and Utility

Applicability and utility refer to the extent to which information presented in a study is relevant for its intended use (U.S EPA, 2019b), in this case informing the evidence assessment and conclusions on atmospheric or exposure science. To be considered for inclusion in an ISA, atmospheric science studies should evaluate the pollutant in ambient air, and in other media as relevant. Exposure studies should measure or estimate exposures that are pertinent to U.S. populations. Applicable studies focus on the policy-relevant issues that help frame the assessment of the health and welfare effects evidence. These issues can vary across ISAs and generally include characterizing emissions sources, transport and transformation processes, measurement and modeling methods, temporal and spatial patterns of ambient air pollutant concentrations and/or exposures, meteorological impacts on pollutant concentrations, relationships and correlations between pollutants within complex mixtures, infiltration to indoor environments, transfers to other media, and the strengths and limitations of various concentration or exposure estimation approaches. For studies examining exposure surrogates, the ISAs evaluate various aspects of exposure measurement error, some of which are discussed below in the context of the health evidence (e.g., errors in measurements; inherent assumptions in exposure modeling; uncertainties in model input parameters; and spatiotemporal variations in pollutant concentrations, activity patterns, and building ventilation). Studies examining variation in pollutant exposures, and the factors contributing to exposure differences between populations, can be particularly informative. The assessment of applicability and utility is aided by clear documentation of study design, data collection/generation techniques, and any underlying assumptions.

[10] Analogous study quality evaluation guidelines have not previously been developed for atmospheric science, and thus the atmospheric science sections of ISAs adapt the study attributes described in the EPA exposure assessment guidelines (U.S EPA, 2019b).

4.1.2. Soundness

Soundness refers to the extent to which the scientific and technical procedures, measures, methods, or models employed to generate information are reasonable for, and consistent with, the intended application (U.S. EPA, 2019b). In evaluating soundness, the ISAs focus on atmospheric science and exposure studies that use quality-assured measurement and/or modeling techniques. Studies should include or reference clear and comprehensive descriptions of the measurement or modeling techniques used, evaluation procedures and performance metrics, strengths and limitations of techniques, and quality-control procedures. In the ISAs, more weight is placed on studies that use methodologies appropriate for meeting study goals and that clearly justify those methodological decisions. Considerations when evaluating the soundness of measurement- and model-based atmospheric science or exposure studies are described below.

- **Measurement studies**: The ISAs place more weight on studies that document or reference high quality measurement data, including uncertainty, bias, sensitivity, specificity, stability, repeatability, and data management practices consistent with best practices among similar measurement studies. Integrated Science Assessments may also consider studies employing innovative, newly developed methods with less established data quality documentation if those methods show promise as potential alternatives, with advantages over established methods, or if they complement established measurement methods by contributing unique insights into atmospheric processes, concentration trends, or exposures.
- **Modeling studies**: The ISAs place more weight on studies that include or reference clear and appropriate rationales for model selection and model assumptions, model input parameters (e.g., the data source for input parameters, the rationale for input data selection, quality of input data in terms of accuracy, precision, representativeness, completeness, and consistency), model calibration and evaluation procedures (e.g., goodness-of-fit criteria for acceptance of the parameter value, procedures to handle outliers, procedures to validate results, model parameter sensitivity analysis, impact of parameter uncertainty on results), and model evaluation results, as well as clear statements of data transfer, transformation and storage procedures.

4.1.3. Clarity and Completeness

Clarity and completeness refer to study transparency, which is a crosscutting attribute essential for evaluating all other attributes of data and study quality (WHO,

2008). For example, sufficient documentation of key assumptions, methods, formulas, input parameters, reasoning processes, and limitations increases the transparency of a study. Key study attributes evaluated for transparency are summarized below.

- **Study Rationale**: Study transparency is greater for studies that present clear rationales for study design, location, population selection (in the case of exposure studies), method selection, and conclusions.
- **Assumptions**: The most informative studies clearly articulate assumptions related to study design and methods used (e.g., dispersion model assumptions, assumptions associated with using central monitoring sites for exposure assignment), as well as assumptions associated with chemistry, transport, dispersion, and/or exposure modeling techniques, and the appropriateness of the assumptions.
- **Data and procedures**: Studies that clearly describe QA and QC procedures, key data, data sources and methods, and statistical methods (e.g., methods for missing data and imputation; quantitative relationships between and within pollutant measurements, such as regression slopes, intercepts, and fit statistic) are considered most informative.

4.1.4. Uncertainty and Variability

Uncertainty comes from incomplete or incorrect information about an environmental measurement or true exposure, and variability describes the natural heterogeneity of measurements or estimates. Uncertainty and variability are associated with each element of both atmospheric and exposure science studies. Evaluation of uncertainty and variability can increase our understanding of the reliability of analyses and the data needed to improve them. In considering studies for inclusion in the ISA, preference is given to studies that characterize uncertainties in measurement and/or modeling approaches and the factors contributing to uncertainties in study analyses. Potential implications of uncertainty and variability in exposure estimates for the health evidence assessment are discussed below in section 4.2.1.4.

4.2. Human Health Studies

Integrated Science Assessments evaluate the quality of health studies using discipline-specific criteria. These criteria are discussed below for epidemiologic (section 4.2.1), controlled human exposure (section 4.2.2), and animal toxicology (section 4.2.3) studies.

4.2.1. Epidemiologic Studies

Epidemiologic studies can examine associations between pollutant exposures and a spectrum of health effects (e.g., changes in heart or lung function, disease development) and outcomes (e.g., hospital admissions, mortality) in the general population and in groups potentially at increased risk. These studies examine the effects of real-world exposures on a variety of populations that can include individuals with severe disease (e.g., asthma, cardiovascular disease) who cannot safely or ethically participate in controlled human exposure studies (see section 4.2.2). Alongside other lines of scientific evidence (i.e., controlled human exposure and animal toxicology studies), epidemiologic studies can provide information supporting conclusions on causality and on other policy-relevant topics including the populations at increased risk, the exposures most closely associated with effects, the lag time between exposure and effect, and the shapes of concentration-response relationships across the range of ambient air pollutant concentrations experienced by populations. In evaluating the quality of epidemiologic studies, ISAs consider the appropriateness of the (1) study design, (2) study population, (3) pollutant(s) evaluated, (4) exposure assessment or assignment, (5) outcome assignment evaluation, (6) control for potential confounding, (7) examination of effect measure modification, and (8) statistical methodology. These study attributes are informed by existing EPA guidelines (U.S. EPA, 2005; U.S EPA, 1998; U.S EPA, 1996; U.S EPA, 1991; U.S EPA, 2019b), and they are discussed below in sections 4.2.1.1 to 4.2.1.8.

4.2.1.1. Study Design

Epidemiologic study designs encompass a variety of study attributes, including the unit of observation, unit of analysis, data source(s), exposure considerations, temporality, randomization, and statistical approach. Broadly, epidemiologic studies that focus on environmental exposures, such as air pollution, can utilize cohort, cross-sectional, time-series, case-crossover, panel, and ecological designs. In addition, quasi-experimental designs have been used to mimic randomized experiments and reduce potential bias. Studies that employ quasi-experimental designs can include intervention studies, studies of natural experiments, and accountability studies. Across the various designs, studies with larger sample sizes increase statistical power and precision and

therefore offer notable advantages over those with smaller sample sizes. National and regional multicity studies that examine associations with health effects across geographical locations are typically conducted in larger and broader populations than those examined in studies of single cities or communities. Multicity studies are also less prone to publication bias than single-city studies (Bell et al., 2005). Other factors being equal, the ISAs generally emphasize results from national- and regional-scale multicity studies when available, noting that they have broader generalizability and can reduce concern for selection bias and publication bias.

Epidemiologic studies examining short-term exposures (e.g., up to 30 days) often employ time-series, case-crossover, or panel designs to evaluate relationships with specific health outcomes at the population level (e.g., mortality, hospital admissions, emergency department visits, symptoms, clinical outcomes, or biomarkers). Time-series and case-crossover studies relate temporal variation in the exposure and outcome, and allow individuals or populations to serve as their own controls, making them less prone than cross-sectional studies to confounding by factors that differ between individuals (e.g., socioeconomic status, age, smoking status). Therefore, among epidemiologic studies of short-term exposures, time-series and case-crossover designs are emphasized over cross-sectional studies. Inference from time-series and case-crossover designs is stronger when they include appropriate controls for factors that vary temporally and that are associated with the exposure of interest and the health outcome (i.e., time-varying confounders; see section 4.2.1.6). Time-series and case-crossover designs share equivalent estimating functions, though time-series studies control for time-varying confounders (e.g., meteorological variables, calendar time, seasonal illness rates) with regression terms, whereas case-crossover studies use a combination of referent matching (e.g., by day of the week, season) and modeling (e.g., meteorological variables).

Panel studies are longitudinal studies that typically examine short-term exposures in smaller study populations. Panel studies can assess outcomes that are not typically examined in larger time-series or case-crossover studies, such as health-related biomarkers, clinical assessments, and symptoms. Additionally, due to smaller sample sizes, panel studies may incorporate personal measures of exposure. Panel studies can

be particularly informative if they employ clearly defined scripted activity patterns (e.g., study participants walk the same routes at the same times of day), measure personal exposures to ambient air pollutants, and measure outcomes at consistent, well-defined lags after exposures.

Epidemiologic studies examining long-term exposures (e.g., longer than 30 days) often employ cohort, cross-sectional, or ecological designs. Epidemiologic studies using a cohort design follow a group of individuals over time while evaluating health outcome(s), which allows for the establishment of temporality, where exposure precedes the outcome, thus strengthening potential causal inferences. Prospective cohort designs can include case-control studies nested within a prospective cohort (e.g., for rare diseases). In contrast, cross-sectional studies measure exposures and outcome(s) at a single point in time, leading to uncertainty regarding the temporal relationship between exposure and effect (i.e., reverse causality cannot be ruled out). Thus, compared to cross-sectional studies, prospective cohort studies can better inform the temporality of the relationship between pollutant exposures and effects, and inference is generally stronger from prospective cohort studies than from cross-sectional studies. Other study quality issues related to design include the validity of inference about individuals from aggregated or group-level data (i.e., for ecological designs in which the units of observation are aggregated) and the appropriateness of control groups (i.e., for case-control studies). Regardless of design, inference from epidemiologic studies of long-term exposure is generally stronger for studies that include appropriate controls for factors associated with the exposure and outcome of interest that vary spatially (i.e., across the study population) and for factors that vary temporally (i.e., long-term trends over time) (confounding is discussed further in section 4.2.1.6).

Some epidemiologic studies employ designs and/or statistical approaches that, compared to traditional regression models, use alternative methods to account for confounders and to mimic randomized experiments and reduce potential bias. In the peer-reviewed literature, these epidemiologic studies are often referred to as causal inference studies, studies that use causal modeling methods, or quasi-experimental studies. As described further in section 4.2.1.6, when evaluating these studies the ISA adopts an approach consistent with that used for evaluating studies that employ

traditional regression models. This approach considers the clarity, plausibility, internal consistency, and validity of study assumptions; the degree to which confounding has been appropriately considered and addressed; and the degree to which statistical uncertainties are appropriately characterized and quantified.

4.2.1.2. Study Population

The population evaluated in an epidemiologic study can impact the strength of inference drawn from that study. The bullets below describe the attributes of epidemiologic study populations considered as part of the ISA study quality evaluations.

- **Representativeness:** There is greater confidence in results for study populations that are recruited from, and representative of, the target population. Selection bias can influence results in either direction and may not affect the internal validity of results but rather reduce the generalizability (i.e., external validity) of findings to the target population.
- **Inclusion and exclusion criteria for participants:** Clearly specified criteria for including and excluding subjects and the reporting of baseline information on participants that are lost to follow-up can aid in evaluating potential selection bias.
- **Participation rates:** Studies with high participation and low loss to follow-up over time that is independent of exposure or health status are considered to have low potential for selection bias.
- **Health conditions:** For studies that evaluate populations with underlying health conditions, independent clinical assessment of the health condition is considered the most reliable approach to identifying the study population, though self-report of physician diagnosis could also be considered a reliable approach for some conditions (e.g., respiratory and cardiovascular diseases). Comparison groups with and without an underlying health condition are more informative if groups are from the same source population.

4.2.1.3. Pollutant

The NAAQS are intended to provide requisite protection of public health and welfare against the effects of exposures to criteria air pollutants. To inform decisions on the NAAQS, each ISA focuses on studies that examine the effects of exposures to the particular criteria pollutant(s) under evaluation. Thus, emphasis is placed on epidemiologic studies that examine associations with individual criteria pollutants, or with components of a criteria pollutant that are particularly relevant for reviewing the

NAAQS (e.g., fine particulate matter ($PM_{2.5}$) in the case of the criteria pollutant particulate matter). Studies only reporting associations with undefined mixtures (e.g., diesel exhaust) or their surrogates (e.g., distance to the roadway) are generally not used to inform ISA conclusions on the health evidence.

Studies of mixtures can be informative in an ISA if independent health effect associations with the particular criteria pollutant under evaluation are also presented. Such studies reflect real-world exposures and can provide insight into combinations of effects or potential modification of the criteria pollutant's effect by the broader pollutant mixture. Inference in the context of the ISA is generally strongest from mixture studies that present (1) independent effect estimates for the pollutant of interest that adequately control for or otherwise address the potential for confounding by co-occurring pollutants (discussed further in section 4.2.1.6) and (2) formal analyses examining how co-occurring pollutants and/or the broader pollution mixture may modify or confound the independent effects of the pollutant of interest. In contrast, studies only presenting associations with mixtures (i.e., with no independent effect estimates for the criteria pollutant under evaluation) are typically not relevant in the context of the ISA.

4.2.1.4. Exposure Assessment or Assignment

Because information about ambient air exposures is rarely available for individual study participants, epidemiologic studies often use surrogates for personal exposures. Exposure surrogates commonly used in epidemiologic studies include concentrations measured by ambient air monitors, modeled ambient or outdoor air concentrations, indoor pollutant concentrations, exposures measured with personal samplers or sensors, and biomarker concentrations. Error in these exposure surrogates can contribute to uncertainty in the associations observed by epidemiologic studies. For example, classical error in the exposure surrogate, defined as error scattered about the true exposure and independent of the true exposure, is generally expected to reduce precision and to bias associations between air pollution and health effects towards the null. Berkson error, defined as error scattered around the measured exposure surrogate (e.g., ambient air concentration) and independent of the true exposure, can affect the precision of the effect estimate by inflating or reducing the standard error (SE). In the absence of

confounders, Berkson error does not bias heath effect estimates, though in the presence of confounders it can lead to bias, generally towards null (Wei et al., 2022).

The ISAs emphasize epidemiologic studies with clear justifications for their exposure surrogates (e.g., in terms of capturing spatiotemporal variation in exposures), as well as studies that compare results across multiple valid exposure assessment methods. The bullets below describe attributes of epidemiologic study exposure estimates that are considered as part of the ISA study quality evaluations.

- **Spatiotemporal variability**: Ambient air concentrations of criteria pollutants vary spatially and temporally. Exposure estimates typically have smaller biases and lead to smaller reductions in precision of associations for pollutants with less spatial variation, compared with spatially heterogeneous pollutants (Sheppard et al., 2005). For pollutants with lower spatial variability, exposure measurement error typically causes health effect estimates to be underestimated (Zeger et al., 2000). In these situations, biases and decreases in the precision of the association (i.e., wider 95% confidence intervals (CIs)) tend to be relatively small (Rothman and Greenland, 1998; Zeger et al., 2000). In the ISAs, studies with validated exposure estimation methods that capture spatiotemporal variability appropriate for the study design and location carry greater weight, especially for studies involving pollutants with relatively large spatiotemporal variabilities (e.g., traffic-related pollutants, such as NO_2). Emphasis is placed on exposure estimation models (e.g., land use regression (LUR), ensemble machine learning methods) for which the model evaluation and/or cross-validation values are available. Additionally, exposure estimation models should include the spatial and temporal resolutions of the model. Emphasis is placed on studies that provide rationale for the exposure assessment approach capturing spatiotemporal variability, model specification or misspecification, and smoothness of the concentration surfaces (Richmond-Bryant, 2020). Spatiotemporal variability can be addressed using location or year as fixed effects, autocorrelation to account for clustering by geography or time, robust errors for unmeasured sources of spatial or temporal variability, hybrid modeling approaches of exposure estimation to capture changes over a smaller scale, hierarchical Bayesian spatiotemporal models, or time-activity data (Talbot et al., 2023; Chen et al., 2018; Zou et al., 2004; Zou et al., 2013; Yoo et al., 2021).
- **Exposure duration:** Regardless of the approach to estimating pollutant exposures, inference is stronger when the exposure duration corresponds with the time course for physiological changes in the outcome (e.g., short-term exposures could result in asthma emergency department visits while multi-year

exposures may be appropriate for development of cancer and other chronic diseases).

- **Lag Time:** When examining outcomes associated with short-term exposures, ISA study quality evaluation may focus on specific lags between exposure and outcome based on evidence for the time course of physiologic changes related to the health effect or outcome being analyzed. When the existence of a lag between exposure and outcome is supported, the following hierarchy is used in the process of selecting results from individual studies of short-term exposures (1 to 4 reflects higher to lower priority):

  (1) Distributed lag models;

  (2) Average of multiple days (e.g., 0−2);

  (3) *a priori* lag days selected by the study authors; or

  (4) Expert judgment considering the time course for physiologic changes for the health effect or outcome if a study focuses on only a series of individual lag days.

- **Exposure estimation methods:** Epidemiologic studies often estimate exposures to ambient air pollutants using fixed-site ambient air monitors, remote sensing approaches, modeling, or a combination of inputs from multiple sources. For some pollutants, studies use biomarker concentrations as estimates of exposure (e.g., blood lead concentrations). General ISA considerations for each of these methods are discussed below.
  - **Fixed-site monitors:** Concentrations reported from fixed-site monitors are most informative if they are correlated with personal exposures; if monitors are located close to study subjects; if monitored concentrations within a location are correlated; or if monitored concentrations are combined with time-activity information. Some studies use (low-cost) sensors to estimate pollutant exposures. Sensors can provide relatively high spatial and temporal resolution, though sensor validation and calibration can be a challenge, because sensor readings may be affected or interfered with by co-occurring pollutants, meteorological conditions, and baseline drifting. These issues should be considered when interpreting studies that use sensors.
  - **Personal monitoring:** Personal monitoring characterizes exposures at the individual level and can provide exposure data attributable to both ambient-air and non-ambient-air sources. Measurement errors associated

with personal monitoring should be documented (e.g., instrument error, degree to which personal exposures represent ambient air exposures).

- **Remote sensing:** Remote sensing-based approaches can use data from satellites to estimate ground-level concentrations of some pollutants (e.g., $PM_{2.5}$, NO_2, and O_3). Remote sensing approaches should be calibrated using ground-level monitoring networks and their accuracy, precision, and any interferences should be clearly documented.
- **Models:** Some epidemiologic studies use atmospheric models to estimate exposures, either in place of or to supplement measurements from ambient air monitors. These models may include LUR models, chemical transport models (e.g., the EPA's Community Multiscale Air Quality (CMAQ) modeling system), dispersion models, geostatistical models (e.g., kriging), population stochastic models,[11] ensemble machine learning models, and hybrid models that use inputs from multiple sources of information (e.g., including models, satellites, and monitors). Uncertainty in exposure predictions from these models is largely influenced by model formulations and the quality of model input data pertaining to precursor emissions or meteorology, which tends to vary on a study-by-study basis. Emphasis is generally placed on exposure predictions from models that incorporate exposure measurements. Studies that use models to estimate exposures should clearly document the procedures for model development and evaluation, as well as model performance under the conditions of the study.
- **Biomarkers:** For some pollutants, epidemiologic studies use biomarkers as surrogate indicators of exposure (e.g., blood lead measured in biomonitoring studies). As noted above for personal monitoring, biomarkers provide exposure estimates at the individual level that can be attributable to both ambient-air and non-ambient-air sources. Some exposure biomarkers may be limited with regard to identifying the specific timing, duration, and/or frequency of the exposure represented, as well as how that biomarker may relate to ambient air concentrations. Thus, depending on the biomarker used there may be limitations in the conclusions that can be drawn about the relationship between the exposure and outcome. When used, biomarkers should be clearly justified

[11] Population stochastic microenvironmental models are not typically used for exposure assessment in epidemiologic studies because the stochastic component of the model can add measurement error to the health effect estimate. However, microenvironmental exposure models can appropriately inform risk and exposure assessments conducted as part of a NAAQS review.

and measured using valid, reliable methods with appropriate characterization of variability.

4.2.1.5. Outcome Assessment and Evaluation

Confidence in study results is greater when outcomes are based on independent clinical assessments of the health condition, without knowledge of exposure status. This can include data from clinical examinations conducted as part of the study, as well as data obtained from hospitals, insurance providers or other health care organizations. Outcomes based on interviews, self-reports, or analysis of biological indicators are generally viewed with relatively high confidence when they are defined by consistent criteria and when outcome data are collected by validated, reliable methods without knowledge of exposure status. For example, independent, clinical assessment is valuable for outcomes like lung function or incidence of disease, but self-report of physician diagnosis may be adequate for outcomes for which validation studies have demonstrated good reliability of self-reported health conditions (Murgia et al., 2014; Weakley et al., 2013; Barr et al., 2002; Muhajarine et al., 1997; Toren et al., 1993). For biological samples, the stability of the biomarker(s) of interest and the sensitivity and precision of the analytical method are considered. In general, if errors in the outcome assessment are not correlated with exposure status, those errors tend to bias results toward the null.

4.2.1.6. Potential Confounding

Confounding is "…a confusion of effects. Specifically, the apparent effect of the exposure of interest is distorted because the effect of an extraneous factor is mistaken for or mixed with the actual exposure effect (which may be null)" (Rothman and Greenland, 1998). A confounder is associated with both the exposure and the outcome. Factors are considered to be potential confounders if they are associated with, but not caused by, the exposure and are risk factors for the health outcome. Not accounting for confounders can introduce bias and may obscure associations or produce artifactual associations. Thus, ISAs evaluate the approaches used in individual epidemiologic studies to account for potential confounding variables and how the strengths and limitations of those approaches might influence weight-of-evidence considerations in causality determinations (NASEM, 2022, pp. 5-6). Inference is stronger for epidemiologic

studies that explicitly identify confounders and their potential impacts on results and that take steps to minimize those impacts, often through statistical adjustment and/or study design. In assessing the evidence for potential confounders, ISAs place greater emphasis on studies that clearly document the appropriateness of models selected, the validity of methods employed, and the assumptions underlying those methods. Studies that provide only correlations or unadjusted (e.g., crude, univariate) effect estimates are not considered for inclusion in ISAs.

Potential confounders vary by study population, exposure, and outcome. To control for confounders, a variety of statistical methods and study design approaches can be employed. Potential confounders included in statistical analyses should be based on a thoughtful review of published literature and the evidence of relationships between variables. In considering the issue of confounding on study conclusions, studies in the ISA may utilize various approaches for identifying potential confounders, controlling for the role of such confounders, and accounting for unknown confounders, as described below.

- **Identifying potential confounders**: Strategies for identifying potential confounders can include *a priori* biological considerations, published literature, causal diagrams (e.g., directed acyclic graphs), and/or statistical analyses.
- **Controlling for confounders:** Scientific judgment is needed to identify the likely sources and extent of confounding and to determine how effectively selected study designs and analyses control for confounders. The approach selected for confounder control should seek to minimize the impact of confounders on reported associations while avoiding overcontrolling (e.g., by adjusting for factors on the causal pathway) and post-treatment confounding caused by inappropriate adjustment for post-treatment variables (e.g., collider bias). Multivariate regression models are often used to detect and control for potential confounders by adjusting for factors that might confound results. These models attempt to control for characteristics that may differ systematically with exposures across study participants. Other approaches that have been applied include, but are not limited to, matching and weighting methods (Stuart, 2010); g-computation and substitution estimators (Keil et al., 2020); doubly robust estimators (Díaz, 2019); bounding approaches (Richardson et al., 2014); quantitative bias analysis approaches (Lash et al., 2014, Weuve et al., 2018); and sensitivity analyses. As described further below, alternative methods for confounder control (i.e., causal

inference frameworks) have also been used in some epidemiologic studies to mimic key aspects of randomized trials.

- **Accounting for unknown confounders:** Approaches to handling unknown confounders include, but are not limited to, instrumental variables (Greenland, 2000); bounding approaches (Richardson et al., 2014); quantitative bias analysis approaches (Lash et al., 2014, Weuve et al., 2018); and sensitivity analyses (e.g., VanderWeele et al. 2017). Additionally, study designs in which individuals or populations serve as their own controls (see section 4.2.1.1) inherently reduce potential confounding by some unknown or unmeasured confounders. In prioritizing studies for inclusion in the ISA, preference is given to studies with clear discussions of the assumptions underlying selected approaches, the robustness of those assumptions, and any sensitivity analyses conducted. Confidence that unmeasured confounders are not responsible for study findings is increased when multiple studies are conducted in various settings using different subjects or exposures and different approaches to confounder control, each of which might eliminate different sources of confounding. Multicity studies can provide insight into the potential impact of unknown confounders on study results through the use of a consistent method to analyze data across locations with different concentrations of copollutants and other covariates.

Some epidemiologic studies employ alternative methods for confounder control that, compared to traditional regression models, are intended to better account for confounders and to mimic randomized experiments and reduce potential bias. These studies are often referred to in the peer-reviewed literature as causal inference studies or studies that use causal modeling methods. In its review of the ISA weight-of-evidence framework, the NASEM committee noted that "the utility of a causal inference framework in an individual study depends on its appropriate use" and "[t]he ability to show mathematical equivalence between statistical and causal parameters does not make the assumptions required for such equivalence true" (NASEM, 2022). The NASEM further noted that well-designed causal inference studies 1) articulate the scientific question in terms of potential and counterfactual outcomes, 2) specify available data and a causal model, 3) articulate a set of assumptions on the causal model that allow the identification of the causal parameter of interest as an observable statistical quantity, 4) analyze the data to estimate the identified statistical quantity, and 5) interpret the statistical results as causal relations according to the validity of the

assumptions made in steps 3 and 4, and quantify the statistical uncertainties (NASEM, 2022, pp. 163-164). The ISA's approach to evaluating studies that use alternative (i.e., causal modeling) methods for confounder control is consistent with the approach to evaluating epidemiologic studies that employ traditional regression models. That is, the ISA considers whether study assumptions are clear and plausible, whether confounding has been carefully considered and addressed with appropriate approaches, whether study assumptions are internally consistent and valid, and whether statistical uncertainties are appropriately quantified (Pearce et al., 2019).

The importance of particular confounders can differ depending on whether short-term or long-term pollutant exposures are evaluated. For studies of short-term exposure, concern is greatest for potential confounders that vary temporally on time scales similar to the variation in exposures and health outcomes. Confounders commonly considered in studies of short-term exposures include, but are not limited to, the following:

- Meteorology,
- Copollutants, particularly those arising from the same source(s) as the pollutant of interest,
- Day of week and season,
- Stress and noise,
- Allergen exposure, and
- Long-term temporal trends.

For studies of long-term exposures, there is greater concern for potential confounders that vary spatially. Spatial confounders commonly considered in studies of long-term exposures include, but are not limited to, the following:

- Socioeconomic status, race/ethnicity, and age,
- Smoking rates,
- Stress and noise,
- Residential housing age,
- Occupational exposures, and
- Short-term exposures.

Studies of long-term exposure that do not also appropriately address relevant time-varying confounders (e.g., societal patterns and trends in smoking rates, medication use) can lead to erroneous conclusions regarding support for causal relationships. A number of methods have been employed to handle time-varying confounders in studies of long-term exposure, including marginal structural models (Robins et al., 2000); regression adjustment using the longitudinal g-computation formula (Bang and Robins, 2005; Hernán and Robins, 2016); longitudinal inverse probability weighting (Ertefaie and Stephens, 2010); doubly robust sequential regression estimators (Díaz et al., 2023; Stitelman et al., 2012); and difference-in-differences designs (Wing et al., 2018; Greenstone and Gayer, 2009).

Across exposure durations, ISAs evaluate the degree to which co-occurring ambient air pollutants may confound health effect associations with the criteria pollutant(s) under evaluation. An emphasis on epidemiologic studies that attempt to minimize such potential "copollutant" confounding may increase confidence in conclusions regarding outcomes associated with the pollutant under evaluation. Copollutant modeling (i.e., two-pollutant models) can reduce concern for this confounding, particularly when correlations between copollutants are relatively low (e.g., $r < 0.4$). However, when correlations are high (e.g., $r > 0.7$), collinearity between pollutants makes copollutant modeling less informative, leading to greater uncertainty in the degree to which reported associations reflect health effects of exposure to the specific criteria pollutant(s) under evaluation.

4.2.1.7. Effect Measure Modification

Effect measure modification occurs when the effect of a pollutant exposure on a health outcome of interest differs between subgroups or strata of risk factors (Rothman and Greenland, 1998). When a risk factor is an effect modifier, it changes the magnitude of the association between the pollutant exposure and the health outcome in stratified analyses. For example, the presence of a pre-existing disease may act as an effect modifier if it is associated with greater risk of effects from air pollution exposure. It is often possible to stratify the association between health outcome and exposure by one or more of these potential effect modifiers. For risk factors that modify the association, effect estimates for each stratum will differ from one another and from the overall

estimate, indicating there to be different quantitative relationships between the exposure metric and outcome for populations represented by these variables.

Inference can be particularly strong from studies that consider the potential impacts of effect measure modification, especially when the modifying factors are coherent with information from other lines of evidence regarding the biological pathways connecting pollutant exposures with particular health effects. For example, larger pollutant-related cardiovascular risks in populations with pre-existing cardiovascular disease could be coherent with animal evidence of cardiovascular effects and controlled human exposure evidence of changes in cardiovascular biomarkers. Studies examining effect measure modification can also be important for quantitatively assessing risks to populations and lifestages at increased risk of health effects for the criteria pollutant evaluated (section 5.3). Inference is stronger from studies that articulate and justify assumptions of treatment effect heterogeneity and that include appropriate diagnostics to account for potentially spurious findings (e.g., with multiple comparisons).

4.2.1.8. Statistical Methodology

Statistical assumptions should be articulated in the context of a study design description. Statistical methods that are appropriate for the power of the study carry greater weight. For example, categorical analyses with small sample sizes can bias results toward or away from the null. Statistical tests such as correlations, t-tests, and chi-squared tests are not considered sensitive enough for adequate inferences regarding health effect associations in epidemiologic studies. Appropriateness and limitations in statistical approaches should be clearly articulated. Sensitivity analyses with alternative statistical models or specifications can inform the stability of findings and aid in judgments of the strength of inference that can be drawn from results. In the case of multiple comparisons, consistency in the pattern of association can increase confidence that associations are not due to chance alone. For all methods, the pattern of effect estimates and precision of the estimates (e.g., indicated by the width of 95% CI) across studies are important considerations for assessing the strength and/or patterns of associations, rather than relying only on statistical significance, which is highly

dependent on study design (Gelman and Greenland, 2019; Greenland et al., 2016; Wasserstein and Lazar, 2016; Wasserstein et al., 2019).

4.2.2. Controlled Human Exposure Studies

Controlled human exposure studies (also known as human clinical studies) evaluate the health effects of experimental exposures in human volunteers under highly controlled and carefully regulated environmental conditions and activity levels. These studies can provide direct evidence of physiological and/or biomolecular effects following air pollution exposures and help to identify the biological pathways linking exposures to health effects in humans. They can provide strong support for the biological plausibility of relationships between air pollutant exposures and health effects that are reported by epidemiologic studies, and precise information on exposure- or dose-response relationships in populations matched for individual level characteristics, often at exposure concentrations at or near those common in ambient air. Thus, data from controlled human exposure studies can provide direct evidence of causal relationships in humans and can help compensate for some of the limitations in epidemiologic studies (e.g., potential confounding by co-occurring pollutants or other factors, like exposure error) (NASEM, 2017). For ethical and practical reasons, controlled human exposure studies are generally limited to examining relatively healthy people or those with mild or moderate diseases; short exposure durations; and exposure concentrations expected to elicit no more than mild, transient effects. These studies generally do not include individuals who may be at the highest risk of pollutant-related health effects (e.g., children, older adults, people with more severe disease or comorbidities).

Controlled human exposure studies should clearly describe the primary and any secondary aims of the study, or specific hypotheses being tested. In evaluating the quality of these studies, the ISAs consider the appropriateness of the study design, study population, pollutants examined, approach to assigning exposures, outcome assessment, control for potential confounders and statistical analysis. Each of these study attributes is discussed below.

4.2.2.1. Study Design

In prioritizing studies for inclusion in the ISA, preference is given to balanced crossover (repeated measures) or parallel study designs that include control exposures to clean air. In a crossover design, each study participant is exposed to both the air pollutant under evaluation and to clean air as a control, under the same conditions (e.g., ventilation rate). In this design, each study participant serves as their own control, minimizing inter-individual confounders. In crossover studies, a sufficient and specified time between exposure days should be provided to avoid carry-over effects from prior exposure days. In studies using a parallel design, all arms should be matched for individual characteristics, such as age, sex, race/ethnicity, anthropometric properties, and health status. In studies evaluating effects of disease, appropriately matched healthy controls can aid interpretation of study results. The evaluation of study design generally includes consideration of factors that can minimize bias in results, such as randomization; blinding; allocation concealment of study subjects, investigators, and research staff; and documentation of withdrawal/exclusion of subjects. Studies should include appropriate control groups to allow for accurate interpretation of results relative to criteria pollutant exposure.

4.2.2.2. Study Population

Depending on the study design, participants recruited into study groups should be matched for age, sex, race/ethnicity, anthropometric properties, and health status. In studies evaluating effects of specific characteristics (e.g., disease, genetic polymorphism), appropriately matched healthy controls are preferred. Relevant characteristics and health status should be reported for each exposure group. For the examination of populations with an underlying health condition (e.g., asthma), independent, clinical assessment of the health condition is ideal, but self-report of physician diagnosis generally is considered reliable for respiratory and cardiovascular disease outcomes. Criteria for including and excluding participants should be indicated clearly, and the loss or withdrawal of recruited participants during a study should be reported.

4.2.2.3. Pollutant

As described above for epidemiologic studies, each ISA focuses on studies that examine the effects of exposures to the criteria pollutant(s) under evaluation. Thus, emphasis is placed on controlled human exposure studies that examine individual criteria pollutants or components of a criteria pollutant that are particularly relevant for reviewing the NAAQS (e.g., $PM_{2.5}$ in the case of particulate matter). Studies of pollutant mixtures (e.g., O_3 as part of an oxidant mixture) can be informative if health effects of exposure to the criteria pollutant under evaluation are also examined separately. Such studies can provide insight into potential modification of the effect of the criteria pollutant by other individual pollutants or by a broader pollutant mixture. Ideally, studies should report the source, purity, and form of the pollutant(s) examined.

4.2.2.4. Exposures

Controlled human exposure studies that approximate expected population exposures in terms of concentration, duration, exertion level, ventilation rate, and method of exposure are of particular interest for the ISAs. In prioritizing studies for inclusion in an ISA, preference is given to studies that evaluate pollutant exposure concentrations close to existing ambient air concentrations in the United States, though studies that use higher exposure concentrations may still provide information relevant to consideration of the dosimetry and/or biological plausibility of effects associated with lower exposures. Studies should have measures in place to adequately monitor and control exposure conditions, including pollutant concentrations, temperature, and relative humidity. The method of exposure (e.g., chamber, facemask) should be specified, and activity level of subjects during exposures should be well-characterized. Preference is given to studies that include control exposures (e.g., to filtered air or room air). Study participants should be randomly exposed without knowledge of the exposure condition, and exposure metrics should be well characterized, including external exposure, intake dose, dosing regimen, and exposure route.

4.2.2.5. Outcome Assessment and Evaluation

For each experiment and each experimental group, including controls, precise details should be provided describing the endpoints examined, how they are measured,

and when and where they are evaluated. Endpoints should be assessed in the same manner for control and exposure groups using valid, reliable methods. This includes using the same procedures in terms of time after exposure, methods, endpoint evaluator, and other attributes. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints such as histopathology. Time of the endpoint evaluations is a key consideration that will vary depending on the endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions, and those time points should be clearly justified.

4.2.2.6. Potential Confounding

To limit potential confounding, studies included in the ISA use either a crossover repeated measures design, in which each study participant serves as their own control, or a parallel exposure design, in which experimental and control groups are matched for individual level characteristics (e.g., race/ethnicity, sex, body weight, smoking history, age) and time-varying factors (e.g., seasonal and diurnal). Exposures should be well-characterized to evaluate independent effects of the pollutant(s) under study.

4.2.2.7. Statistical Methodology

Statistical methods should be described clearly and be appropriate for the study design and research question, including correction for multiple comparisons when appropriate. Statistical significance is generally used to evaluate the findings of individual controlled human exposure studies. Detection of statistical significance is influenced by a variety of factors, including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion, though inference is strongest for studies with sample sizes that provide adequate power to detect hypothesized effects. Because statistical tests have limitations, consideration is also given to trends in data and reproducibility of results. Consistent trends across studies can be informative, even if results of individual studies are not statistically significant.

4.2.3. Animal Toxicology Studies and New Approach Methodologies

Animal toxicological studies evaluate the health effects of air pollutants using animal models (e.g., non-human primates, mice, rats, and guinea pigs) under well-

controlled pollutant exposure environments. Investigators expose non-human mammalian animal species to known concentrations of air pollutants under carefully regulated laboratory conditions. Experimental animal studies provide critical information on potential human health effects, exposure- and dose-response relationships, and underlying toxicological pathways and mechanisms of action. One major uncertainty in animal studies is the extent to which animal responses predict human outcomes, given potential differences in metabolism, hormonal regulation, breathing patterns, lung structure, physiology, and anatomy. When these differences are appropriately characterized, experimental animal studies can inform and improve our understanding of biological mechanisms, providing support for the biological plausibility of relationships between air pollutant exposures and health effects that may be indicated by epidemiologic studies, and can address uncertainties in other lines of evidence (e.g., confounding in epidemiologic studies). In addition, depending on the animal model used and the study design employed, experimental animal studies may help inform an understanding of the potential for biological responses and health effects in certain populations that may be at increased risk of a criteria pollutant-related health effect.

Emerging approaches in toxicology include new approach methodologies (NAMs), a term which refers to any technology, methodology, approach, or combination of these tools that can inform chemical hazard and risk assessment while avoiding the use of animal testing. This can include *in silico*, *in chemico*, *in vitro*, and *ex vivo* approaches (ECHA, 2016; U.S. EPA, 2018; U.S. EPA, 2021).[12] While ISAs have historically relied on *in vivo* studies to provide mechanistic insight, or *in vitro* models of cancer-related effects (e.g., mutagenesis), data collected using well-validated NAMs may also provide insight into the occurrence of health effects when guidelines are established for the use of these methods in air pollution studies.

[12] Experimentation performed using a computer is referred to as an *in-silico* approach (Cronin, 2009). Chemical reactivity tests that are performed in the absence of biological materials are referred to as *in chemico* approaches (Cronin, 2009). *In vitro* tests are performed outside of a living organism using established cell lines, while *ex vivo* approaches are those that involve the collection of cells/tissues/organs from living organisms for use in experimentation (European Commission et al., 2020).

As for other disciplines, toxicological studies should clearly describe the primary and any secondary aims of the study, and the specific hypotheses being tested. In evaluating the quality of these studies, the ISAs consider the appropriateness of the study design, test model, pollutant(s) examined, exposure assignment and approach, outcome assessment, variable control, and statistical analysis. Each of these study attributes is discussed below.

4.2.3.1. Study Design

Toxicological studies should include appropriately time-matched control exposures, should randomize assignment to exposure groups and, where possible, should blind research personnel from endpoint evaluation and analysis. Blinding of research personnel to study groups may not be possible due to animal welfare and experimental considerations; however, differences in the monitoring or handling of animals across groups by research personnel should be minimized. Studies should use methods to limit differences in baseline characteristics of control and exposure groups, and groups should be subjected to identical experimental procedures and conditions to the extent possible (e.g., in terms of animal care and housing). The evaluation of study design generally includes consideration of factors that can minimize bias in results, such as randomization, blinding, and documenting any loss of animals. Where applicable, approval of study protocols by appropriate institutional animal care and use committees must be obtained (European Commission et al., 2020; Cronin et al., 2009).[13]

4.2.3.2. Test Model

Toxicological studies should provide a clear justification for their chosen model system. Unless data indicate otherwise, laboratory nonhuman mammalian animal species of any lifestage, stock, and strain are considered appropriate for evaluating the effects of pollutant exposure. Ideally, studies should report species, strain, substrain, genetic background, age, sex, and weight. It is preferred that the authors test for effects in both sexes and multiple lifestages and report results for each group separately. All

[13] For historical studies that pre-date institutional animal care and use committee requirements detailed in the Animal Welfare Act, a modified approach to study quality evaluation is applied where use of an animal care and use committee is not an absolute requirement.

animals used in a study should be accounted for, and the rationale for any exclusion of animals or data should be provided.

4.2.3.3. Pollutant

As described above for other disciplines, each ISA focuses on studies that examine the effects of exposures to the particular criteria pollutant(s) under evaluation. Thus, emphasis is placed on toxicological studies that examine individual criteria pollutants or components of a criteria pollutant that are particularly relevant for reviewing the NAAQS (e.g., $PM_{2.5}$ in the case of particulate matter). Studies of pollutant mixtures can be informative if the effects of exposure to the criteria pollutant under evaluation are also examined separately. Such studies can provide insight into potential modification of the effect of the criteria pollutant by other individual pollutants or by a broader pollutant mixture. Ideally, studies should report the source, purity, and form of the pollutant(s) examined.

4.2.3.4. Exposures

Animal toxicological studies should have measures in place to adequately control exposure conditions, including the identity of the target pollutant, exposure concentrations, temperature, and relative humidity. Studies that approximate expected human exposures in terms of concentration, duration, timing of exposure, and method of pollutant administration are of particular interest, though higher exposure concentrations may be considered relevant for some pollutants (e.g., within one to two orders of magnitude of recent ambient air concentrations). This range in relevant exposures is meant to account for differences in dosimetry, toxicokinetics, and biological sensitivity between humans, including groups at increased risk, and the various animal species and strains used in toxicological studies. While inference is stronger for studies examining exposure concentrations close to those in ambient air, studies examining exposure concentrations or doses at the higher end of the range may be considered if they provide insight into an understudied effect, the mode of action for an observed effect, an exposure-response relationship encompassing multiple exposure concentrations, inter-species variation, or factors that may confer increased risk in human populations.

The focus for ISAs is generally on inhalation exposure, but oral and intravenous exposures may also be informative for pollutants for which other exposure routes are relevant and/or if a biomarker (e.g., blood lead) is being examined. Non-inhalation exposure experiments that target the respiratory tract (e.g., intratracheal instillation [IT]) may be informative if they provide information relevant to biological plausibility and dosimetry. As discussed above, *in vitro* studies in validated model systems may be considered for inclusion in an ISA, though the relevance of *in vitro* exposure concentrations to human inhalation exposures are often an additional source of uncertainty. All studies should include exposure control groups (e.g., filtered air or room air).

4.2.3.5. Outcome Assessment and Evaluation

For each treatment group, including controls, precise details should be provided describing the endpoints examined, how they are measured, and when and where they are evaluated. Endpoints should be assessed in the same manner for control and exposure groups using valid, reliable methods. This includes using the same procedures in terms of time after exposure, methods, endpoint evaluator, and other attributes. Limits of detection should be provided for quantitative assays, when available. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints such as histopathology. Timing of the endpoint evaluations is a key consideration that will vary depending on the endpoint under investigation. Endpoints should be assessed at time points that are appropriate for the research questions.

4.2.3.6. Variable Control

To limit potential impact of other variables on study results, studies included in the ISA should match experimental and control groups for individual-level characteristics and time-varying factors. Individual characteristics include strain, sex, body weight, litter size, feed, and water consumption. Exposures should be well characterized to evaluate independent effects of the pollutant(s) under study.

4.2.3.7. Statistical Methodology

Statistical methods should be described clearly and should be appropriate for the study design and research question, including correction for multiple comparisons when

appropriate. Results should be reported with sufficient detail to allow for independent interpretation (e.g., quantitatively with a measure of variance). Statistical significance is generally used to evaluate individual study findings. Detection of statistical significance is influenced by a variety of factors, including, but not limited to, the sample size used in the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion, though the sample size should provide adequate power to detect hypothesized effects. Because statistical tests have limitations, the ISAs also consider trends in results across toxicology studies and reproducibility of study results. Consistent trends across studies can be informative for weight-of-evidence evaluations, even if results of individual studies are not statistically significant.

4.3. Ecological and Other Welfare Effects

Integrated Science Assessments define an ecosystem as "a functional unit consisting of living organisms, their nonliving environment, and the interactions within and between them" (IPCC, 2014). Ecosystems, whether natural, cultivated, or urban (U.S. EPA, 1986), are characterized by their structure and function. Ecosystem structure is the physical arrangement of abiotic and biotic components and includes climatic and edaphic components, species abundance, richness, distribution, diversity, evenness, and composition. Ecosystem function refers to the suite of processes and interactions among the ecosystem components that involve energy or matter. Examples include energy flow and biogeochemical cycling.

The ISAs evaluate ecological endpoints across multiple levels of biological organization and spatial scales. Scales of biological organization describe attributes at the suborganismal, organismal, population, community, or ecosystem-level. In ecology, a population refers to a group of individuals of the same species. A community is an assemblage of interacting populations of different species in the same spatial area. Communities, together with the physical environment, constitute ecosystems. Ecosystems do not have fixed boundaries, rather, some ecosystems are linked by gradual, environmental shifts while others transition abruptly. Thus, the extent of an

ecosystem depends on the scale of the study and ranges from very small spatial scales to, ultimately, the biosphere (IPCC, 2014).

Criteria air pollutants can impact terrestrial, freshwater, and coastal ecosystems as these ecosystems are connected through hydrological, biogeochemical, and atmospheric cycles. Exposure may be direct, via gaseous or aerosol forms, or indirect, through deposition to soil, water, vegetation, or fauna. Some air pollutants (e.g., Pb, oxides of nitrogen, sulfur oxides) may undergo subsequent transport and transformation through environmental media (air, soil, water, sediment, and biota), while others (e.g., ozone and $PM_{2.5}$) can be taken up directly by vegetation from the atmosphere via dry deposition. Direct and indirect pollutant exposure can affect ecological endpoints across all biological scales. Initial exposures may induce subcellular and cellular changes in organisms which subsequently impact higher levels of biological organization and ecosystem functions. Uptake of atmospherically deposited criteria air pollutants from the environment, subsequent bioaccumulation, and toxicity vary within and among biological species and taxa.

To accommodate the diversity of organisms, levels of biological organization, endpoints, study designs, analysis methods, modeling approaches, and reporting conventions in welfare effects studies, the criteria for evaluating study quality are aligned with the broad questions provided in the introduction to section 4. Welfare effects studies are reviewed for appropriateness of study design, exposure methodology, statistical analyses, control of confounding factors, and other relevant criteria. Studies are screened to ensure the exposure methods, including assumptions, uncertainties, and limitations, are clearly described, suitable for the research question, and include appropriate QA testing. Studies that expose ecosystems to air concentrations above ambient are included if they also test concentrations relevant to ambient air, inform understanding of modes of action, or illustrate the range of sensitivity to air pollutants across taxa, biomes, or ecoregions. Welfare effects studies are typically excluded from further consideration in the ISAs if exposure concentrations exceed the concentration limits specified in the LECES statement (Table 6), concentrations are not reported, the effect of interest cannot be attributed to a specific pollutant (e.g., in a mixtures study), effects on endpoints of interest are not tested

statistically, description of methods are inadequate or missing, or study design is inadequate.

Evidence is considered from multiple and diverse types of studies including those that best approximate the concentration-, exposure-, or dose-response relationships between welfare endpoints and the pollutant. Controlled experimental studies provide the most direct and quantifiable information on the relationships between pollutant exposures and welfare effects. To the extent available, the ISAs also evaluate results from less controlled field studies that may provide information on the relationship between air pollutant exposure conditions and an endpoint. Other types of studies or data can inform an understanding of concentration-, exposure-, or dose-response relationships, including results from observational studies, modeling studies parameterized with field data, or meta-analyses. These and other types of studies aid in identifying modes of action and characteristics of sensitive ecosystems. Studies may also inform an understanding of specific air quality conditions that increase the risk of effects from pollutant exposures. For example, the evidence base for visibility impairment associated with PM indicates visibility can vary with humidity and PM composition.

In evaluating studies on non-ecological welfare effects, such as climate and visibility impacts, emphasis is placed on the most recent international assessments (i.e., for climate), on studies that use well-established measurement and modeling techniques, and on studies that report uncertainty or compare results from a variety of techniques. Studies that use satellite observations may also be informative in addressing knowledge gaps. Relevant climate studies include those evaluating direct impacts of criteria air pollutants on Earth's energy balance, as well as other physical climate variables such as temperature changes at regional or global scales. Studies that evaluate effects by source sector or region (e.g., regional climate modeling studies) are particularly informative. Studies that report impacts of multiple criteria pollutants (and multiple components of PM) are useful in evaluating interactions and the relative contributions of atmospheric constituents. For example, in evaluating the climate forcing effects of ozone, it is useful to understand the atmospheric chemistry involving the chemical precursors (including CO and nitrogen oxides) that influence its atmospheric concentrations. In the case of visibility, studies conducted in the United States and

Canada provide information relevant for reviewing the NAAQS. Particularly relevant are studies on visibility impairment and valuation studies that explicitly separate rules for visibility[14] from concerns about health risks of air pollution.

5. SYNTHESIS AND INTEGRATION OF EVIDENCE AND DEVELOPMENT OF SCIENTIFIC CONCLUSIONS

After evaluation of individual scientific studies as described above in sections 2 through 4, EPA scientists synthesize and integrate the evidence that is judged to be of sufficient relevance and quality. Integrated Science Assessments draw from this synthesis and integration to develop conclusions on policy-relevant topics, including the strength of evidence supporting causal relationships between criteria pollutant exposures and health or welfare outcomes (i.e., causality determinations). Integrated Science Assessments also include targeted evidence evaluations and conclusions on other policy-relevant topics in order to provide broader insight into the public health or welfare impacts of pollutant exposures and into the strengths and limitations of the evidence supporting those impacts. The other topics evaluated vary across ISAs and outcomes, depending in part on the strength of supporting evidence. Examples of other topics commonly evaluated in ISAs include the following:

- The factors that may confer higher risk for specific populations and/or lifestages;
- The exposure conditions under which effects can occur;
- The strengths and limitations of the available evidence with regard to identification of effects and the exposure conditions expected to elicit effects.

The ISA causality determinations and conclusions on policy-relevant topics are most fundamentally drawn from the synthesis and integration of the scientific evidence described below in section 5.1. The ISA framework for reaching causality determinations based on this synthesis and integration is described in section 5.2. Section 5.3 discusses evaluation of the broader public health impacts of criteria pollutant exposures, and section 5.4 discusses evaluation of the broader ecological and other public welfare impacts of criteria pollutant exposures.

[14] Available at https://www.epa.gov/visibility/visibility-regulatory-actions

5.1. Evidence Synthesis and Integration

The ISA approach to synthesizing and integrating scientific evidence draws from historical efforts to characterize the weight of evidence supporting causation. The 1964 Surgeon General's report on tobacco smoking discussed criteria for the evaluation of epidemiologic studies, focusing on consistency, strength, specificity, temporal relationship, and coherence (Hew, 1964). Sir Austin Bradford Hill (Hill, 1965) articulated similar aspects of causality in epidemiology and public health that have been widely adopted (IOM, 2008; IARC, 2006; U.S. EPA, 2005; CDC, 2004). The EPA has adapted the aspects from these historical efforts for use in reaching ISA causality determinations.

Table 7 (below) and the accompanying text describe the key aspects considered in synthesizing and integrating the scientific evidence. These aspects provide a framework for assessing the evidence, though they do not lend themselves to being considered in terms of simple formulas or fixed rules (Hill, 1965). For example, one cannot simply count the number of studies reporting statistically significant results or statistically nonsignificant results and reach credible conclusions about the relative weight of evidence and the likelihood of causality. The aspects cannot be used as a strict checklist, and failure to meet one or more of the principles does not automatically preclude a determination of causality (CDC, 2004). Rather, these aspects provide a framework for systematic appraisal of the body of evidence, informed by peer review and public comment (see section 6), which includes weighing alternative views on controversial issues.

Table 7. Aspects of the evidence considered in judging causality.

Aspect	Description
Consistency	An inference of causation is strengthened when a pattern of elevated risks is observed across multiple independent studies. The reproducibility of findings in different groups and using different study designs constitutes one of the strongest arguments for causation. If there are discordant results among investigations, possible reasons such as differences in exposure, confounding factors, and the power of the study are considered.
Coherence	An inference of causation is strengthened when multiple lines of evidence independently support the occurrence of related effects following pollutant exposure. Coherence can be demonstrated by evidence across various disciplines and/or study designs.
Experimental evidence	Strong evidence supporting causation can be provided by experimental studies, in which exposures are manipulated by investigators, and by studies of natural experiments when a change in exposure is found to result in a change in the occurrence, frequency, or severity of health or welfare effects.
Biological plausibility	An inference of causation is strengthened by experimental or other studies demonstrating biologically plausible mechanisms through which pollutant exposures could lead to adverse outcomes.
Biological gradient (i.e., exposure-response)	A well-characterized exposure- or dose-response relationship (e.g., larger, more serious effects associated with higher exposure/dose) can strongly support causation, especially when such relationships are observed across multiple studies, disciplines, and durations of exposure.
Strength of the observed association	The finding of large, precise risks can increase confidence that an association is likely not due to chance, bias, or other factors. However, an effect estimate that is small in magnitude does not necessarily indicate a lack of causation.
Specificity of the observed association	Evidence linking exposure to a specific outcome can provide support for causation. However, lack of specificity does not necessarily indicate a lack of causation, since it is rarely expected that a pollutant exposure will invariably predict the occurrence of an outcome, and since a given outcome may have multiple causes.
Temporality of the observed association	Evidence that pollutant exposure precedes the appearance of the effect, and that the interval between exposure and effect are reasonable based on available science, supports causation.

5.1.1. Consistency

In assessing the consistency of findings across studies, ISAs emphasize the pattern of results in a body of studies examining related outcomes. Consistency of findings can be supported by the repeated observation of effects or associations across multiple independent studies. Results that are reproducible across variations in study designs or analytic choices "can be viewed as more robust and as stronger evidence for a causal relationship" (NASEM, 2022, p. 7). Thus, the strength of the evidence is

increased when similar findings are reported in different populations, organisms, or ecosystems under different circumstances.

The evaluation of consistency in an ISA includes consideration of the direction and magnitude of effects across independent studies. For epidemiologic studies and studies of ecological effects, ISAs place greater emphasis on the pattern of results across studies than on the statistical significance of results in individual studies. Statistical inferences may result in both false positives and false negatives, and statistical significance is influenced by a variety of factors including, but not limited to, the size of the study population, exposure and outcome measurement error, and statistical model specifications. Inconsistent results among independent studies may be explained by differences in study methods, random errors, exposure errors, confounding factors, or study power. Thus, while statistical significance in epidemiologic studies and studies of ecological effects may be informative, it is only one means of evaluating confidence in the observed relationship and assessing the probability of chance as an explanation. Statistical significance of results is traditionally given greater emphasis in the evaluation of consistency across controlled human exposure and animal toxicological studies, which are conducted under carefully controlled laboratory conditions, though the pattern of results across such experimental studies using similar designs and examining related endpoints can also be informative.

5.1.2. Coherence

In evaluating coherence of the evidence, ISAs examine the degree to which studies from different disciplines, or studies from the same discipline with different designs, support the occurrence of related effects as a result of pollutant exposures. For example, the health evidence may include epidemiologic studies reporting positive associations between pollutant exposures and cardiovascular events that are coherent with controlled human exposure and/or animal toxicology studies demonstrating changes in cardiac or vascular function following exposures. Evidence that is coherent across disciplines and/or study designs, and across levels of biological organization for studies of ecological effects, provides stronger support for causality than an individual line of evidence alone.

For bodies of evidence characterized by limited coherence, the implications for causality determinations and other conclusions are carefully considered. It is important to account for how study design limitations can impact results and potentially explain inconsistencies in the broader evidence base. For example, caution may be appropriate when reaching conclusions on the absence of health effects, or on the exposure concentrations below which effects no longer occur, based on controlled human exposure studies that examine relatively healthy volunteers with short exposure durations and short outcome evaluation periods. Such study characteristics could bias results towards not detecting effects that may occur in less healthy populations and/or with longer-term exposures or evaluation periods.

5.1.3. Experimental Evidence and Biological Plausibility

Integrated Science Assessments evaluate the extent to which experimental studies, in which exposures are experimentally manipulated by investigators, provide direct evidence of effects resulting from air pollutant exposures. For human health, animal toxicological and controlled human exposure studies provide valuable information on the relationships between exposures and observed effects under well-defined conditions, and they can provide insight into the biological plausibility of causal relationships with outcomes examined in epidemiologic studies. For studies of ecological effects, experimental studies at lower scales of biological organization (e.g. cellular, organismal) are often essential for understanding observations from higher levels of organization (community, ecosystem), where the direct manipulation of exposures and control of other variables is generally not practical.

Support for biological plausibility can come from studies that provide an understanding of the mode of action through which pollutant exposures lead to health or welfare effects. This understanding may span multiple levels of biological organization, including molecular and cellular events in the pathways leading to disease or other adverse effects. A complete understanding of the mode of action is often not available and is not necessary for it to be biologically plausible that an exposure-response relationship reflects a causal relationship. Rather, studies demonstrating that a pollutant exposure elicits key physiological events in the exposure-to-response pathway

can provide strong support for biological plausibility, even in the absence of data on the complete pathway.

Experimental studies that approximate expected human or ecosystem exposures in terms of concentration, duration, timing of exposure, and method of pollutant administration are of particular interest for evaluating biological plausibility. However, higher exposure concentrations in animal models of human health may be considered relevant to account for differences in dosimetry, toxicokinetics, and biological sensitivity between humans, including groups at increased risk, and the various animal species and strains used in toxicological studies.

5.1.4. Biological Gradient

The presence of concentration-, exposure-, and/or dose-response relationships can increase confidence in a conclusion that pollutant exposures can cause observed effects, particularly when such relationships are demonstrated across scientific disciplines and/or in multiple independent studies in which confounders have been adequately controlled. The shapes of concentration-, exposure-, or dose-response curves, and whether those curves are linear across the range of ambient air exposures, can also be an important consideration in characterizing the public health and welfare impacts associated with pollutant exposures. Sources of variability and uncertainty in interpreting these relationships can include limitations in the data available toward the lower and upper ends of the concentration range, the possible influence of exposure measurement error over the range of concentrations, and variability in response among individuals. These sources of variability and uncertainty tend to smooth and "linearize" concentration-, exposure-, and dose-response functions. As a result, they can obscure the existence of nonlinear relationships and thresholds, should they exist.

5.1.5. Strength and Specificity of Associations

In evaluating the strength and specificity of observed associations, the ISAs consider the magnitude and statistical precision (i.e., informed by width of confidence intervals) of the associations, as well as the specific effects reported and the populations impacted, across studies. In large studies that adequately account for potential confounding factors, strong associations can serve to increase confidence that findings

are not due to a weak unmeasured confounder, chance, or other biases. However, effects evaluated in ISAs tend to have multiple contributing factors (e.g., genetics, disease, lifestyle, and environmental factors for health effects), and the magnitude of the contribution from air pollution exposures will depend on the prevalence of other risk factors in the study population. Thus, in studies that appropriately account for potential confounding factors and other sources of bias a small effect size does not rule out a causal relationship with the air pollutant, particularly when the specific effect and the population(s) impacted are consistent across studies. For human health, a small effect can be important from a public health perspective if it impacts large segments of the population. A small effect can represent a shift in the distribution of responses in the study population and may increase the proportion of individuals with clinically important outcomes.

5.1.6. Temporality of the Observed Association

Temporality of an association refers to the temporal sequence of exposure and observed effects. For a causal relationship to exist between air pollution exposure and a health or welfare effect, the exposure must precede the effect. Not all observational studies provide evidence of temporality. For example in epidemiology, cohort studies are generally better suited to address the consideration of the temporal sequence of exposure and effect than are cross-sectional studies (see section 4.2.1.1). Strong support for appropriate temporal relationships between exposures and effects can also be provided by experimental studies.

5.2. ISA Framework for Causality Determinations

The 1964 Surgeon General's report on tobacco smoking defined "cause" as a "significant, effectual relationship between an agent and an associated disorder or disease in the host" (Hew, 1964). More generally, a cause is an agent that brings about an effect or a result. Unlike an association, a causal claim supports the creation of counterfactual claims; that is, a claim about what the world would have been like under different or changed circumstances (IOM, 2008).

Drawing from the evidence evaluation and integration described above in section 5.1, causality determinations reflect the strength of evidence supporting causal

relationships between criteria pollutant exposures and particular effects. Evaluating the support for causal relationships between criteria pollutant exposures and health or welfare effects is made challenging by the fact that many of the effects or outcomes evaluated in ISAs have complex etiologies. Diseases such as asthma, coronary heart disease, and cancer are typically initiated by multiple biological and environmental factors. Effects in an individual can depend on factors such as age, genetic background, nutritional status, immune competence, and social characteristics (IOM, 2008; Gee and Payne-Sturges, 2004). Similarly, welfare outcomes (e.g., related to plant growth, ecosystem health) can be affected by a number of factors that include, but are not limited to, air pollution exposures. The combination of multiple factors could result in synergistic or antagonistic effects, and the observed health and welfare impacts of criteria pollutant exposures may represent the net effect of many actions and counteractions.

Given the complexities inherent in examining these relationships, in its review of the 2015 Preamble to the ISAs NASEM asserted that "[a] single study will rarely definitively and comprehensively address issues associated with the determination of causality that are examined in ISAs" (NASEM, 2022; p. 3). Compared to any single study, the ISA causality framework recognizes that the availability of multiple studies evaluating a particular topic, each with different strengths and limitations, provides a more robust foundation for evaluating the weight of evidence. A weak inference from one line of evidence can be addressed by other lines of evidence. This approach is consistent with the NASEM conclusion that "A weight of evidence approach—combining assessment of study quality with expert judgment—allows EPA to draw conclusions that integrate scientific findings across multiple study designs and disciplines, as required by the CAA" (NASEM, 2022; p. 3).

The description of the ISA causality framework provided in this section builds on the 2015 Preamble (U.S. EPA, 2015), with updates reflecting NASEM recommendations (NASEM, 2022) and advances made over multiple ISAs since 2015. The ISA causality framework has been informed by other weight-of-evidence approaches formulated by regulatory and science organizations, including the National Academy of Sciences (NAS) Institute of Medicine (IOM, 2008), the International Agency for Research on Cancer

(IARC, 2006), the Centers for Disease Control and Prevention (CDC, 2004), the World Health Organization (WHO, 2021), and the EPA (U.S. EPA, 2005; U.S. EPA, 2022). The frameworks used by each of these organizations are similar in nature but adapted to different purposes and have proven effective in providing uniform structure and language for causality determinations.

Using the aspects of the evidence described above to inform judgments, the ISAs evaluate the relevant scientific literature and draw conclusions on the strength of support for causal relationships between relevant pollutant exposures and health and welfare effects. These "causality determinations" reflect overall confidence in the causal nature of relationships based on the strengths and limitations of the full body of evidence, integrated within and across disciplines. In its review of the causality framework in the 2015 Preamble to the ISAs (U.S. EPA, 2015), the NASEM supported this approach, noting that "[a] weight of evidence approach, which combines assessment of the scientific literature with expert judgment to weigh that complex literature, is a scientifically defensible approach for the ISA causal[ity] determination framework" (NASEM, 2022, p. 126).

The ISAs evaluate evidence for major health and welfare effects categories or groups of related endpoints (e.g., respiratory effects), characterizing the strengths and limitations of evidence for individual endpoints within the broader category. Limitations in the evidence base can result from the consistent presence of related uncertainties within a group of studies (e.g., studies similarly affected by confounding, exposure error, and/or species extrapolation) or uncertainties that exist across the broader body of evidence (e.g., inconsistent evidence across disciplines, lack of coherence). Integrated Science Assessments generally rely on qualitative uncertainty evaluations, though quantitative analysis approaches such as meta-regression may be used in some situations.[15] Rigorous external peer-review by the CASAC is critical to informing ISA

[15] Meta-analyses of existing literature may be considered for inclusion in an ISA (section 2.3). Pooled effect estimates from meta-analyses provide precision-weighted quantitative summary estimates that account for random error, though these pooled estimates are considered with caution when evaluating causality, so as not to misrepresent and perpetuate errors resulting from consistent biases across these study attributes (Savitz and Forastiere, 2021, Savitz et al., 2019).

conclusions on the strengths, uncertainties, and/or bias in the body of evidence supporting causality determinations (see section 6).

The ISA causality determinations are articulated using a framework with a five-level hierarchy characterizing the weight of evidence for causation (Table 8). The NASEM endorsed this approach, noting that the five categories for classifying causality determinations "are scientifically defensible given the precautionary nature of the CAA" (NASEM, 2022, p. 3). The standardized language used in the framework to describe specific determinations is adapted from sources across the federal government and the wider scientific community, especially the EPA *Guidelines for Carcinogen Risk Assessment* (U.S. EPA, 2005), U.S. Surgeon General's report, *The Health Consequences of Smoking* (CDC, 2004), and NAS Institute of Medicine document, *Improving the Presumptive Disability Decision-Making Process for Veterans* (IOM, 2008). Table 8 presents the human health and welfare descriptors for each of the determinations in the ISA causality framework.

Table 8. Causality framework for health and welfare effects.

Descriptor	Health Evidence Characteristics	Ecological and Other Welfare Evidence Characteristics
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures. That is, pollutant exposures have been shown to result in health effects across studies in which chance, confounding, and other biases can be ruled out with reasonable confidence. Evidence that is sufficient to infer a "causal" relationship generally includes multiple high-quality studies conducted by different research groups and this determination draws from the joint consideration of several lines of evidence that reinforce each other. Evidence supporting this determination can include controlled human exposure studies that consistently demonstrate related effects and/or observational epidemiologic studies reporting consistent health effect associations that, when considered in light of study quality and coherence with other lines of evidence (i.e., controlled human exposure	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures. That is, the pollutant has been shown to result in effects across studies in which chance, confounding, and other biases can be ruled out with reasonable confidence. Controlled exposure studies (laboratory or small- to medium-scale field studies) provide the strongest evidence for causality, but their scope of inference may be limited. Generally, this determination is based on multiple studies conducted by different research groups, and evidence that is considered sufficient to infer a causal relationship is usually obtained from the joint consideration of many lines of evidence that reinforce each other.

	studies, animal toxicological studies, mode of action information), are not explained by plausible alternatives.	
Likely to be a causal relationship	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures. That is, pollutant exposures have been shown to result in health effects in studies where chance, confounding, and other biases are minimized, but uncertainties remain in the evidence overall. A "likely to be causal" relationship is generally based on multiple high-quality studies conducted by different research groups. Evidence supporting this determination can include 1) multiple high-quality observational epidemiologic studies consistently reporting health effect associations, but with uncertainty remaining related to potential confounding or coherence with other lines of evidence (i.e., controlled human exposure studies, animal toxicological studies, mode of action information); 2) consistent evidence in animal models and/or other experimental models (e.g., for cancer-related effects) that can be reasonably extrapolated to human health, but only limited human data indicating an effect.	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures. That is, an association has been observed between the pollutant and the outcome in studies in which chance, confounding, and other biases are minimized but uncertainties remain in the evidence overall. For example, field studies show a relationship, but suspected interacting factors cannot be controlled, and other lines of evidence are limited or inconsistent. Generally, the determination is based on multiple studies by different research groups.
Suggestive of, but not sufficient to infer, a causal relationship	Evidence is suggestive of, but not sufficient to infer, a causal relationship with relevant pollutant exposures. That is, pollutant exposures have been associated with health effects in some studies, but chance, confounding, and bias cannot be ruled out with confidence. Evidence supporting a "suggestive" relationship can include studies of varying quality that may be generally supportive of pollutant-related effects, but not entirely consistent and with limited coherence across lines of evidence. A suggestive determination can be reached with relatively small bodies of evidence or, in rare cases, one high quality study.	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but chance, confounding, and bias cannot be ruled out with confidence. For example, at least one high-quality study shows an effect, but the results of other studies are inconsistent.
Inadequate to infer the presence or absence of a causal relationship	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. That is, supporting evidence is limited and available studies are of insufficient quantity, quality, consistency, and/or statistical power to permit a conclusion	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quantity, quality, consistency, and/or statistical power to permit a conclusion regarding the presence or absence of an effect.

	regarding the presence or absence of an effect.	
Not likely to be a causal relationship	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies, covering the range of relevant exposures and considering at-risk populations and lifestages, are consistent in not showing an effect at any level of exposure.	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies examining relationships with relevant exposures are consistent in not showing an effect at any level of exposure.

Each level of the causality hierarchy is delineated by the degree to which chance, confounding, and other biases can be ruled out as explanations of reported effects with reasonable confidence. The conclusion on which level within the hierarchy best fits a particular body of evidence is informed by considering the aspects of the evidence described above in section 5.1 (i.e., consistency, coherence, biological plausibility and availability of experimental evidence, biological gradient, strength and specificity of associations, and temporality). Integrated Science Assessments use these aspects to guide evidence integration both across populations and within groups that share characteristics that may place them at higher risk. This approach is consistent with the NASEM caution that "considering, or highlighting, only overall average population or broad ecosystem effects can obscure causal relationships that exist for more sensitive subgroups, subspecies, communities, or ecosystems" (NASEM, 2022, p. 128). Thus, as recommended by NASEM (NASEM, 2022, p. 4), when studies consistently demonstrate elevated pollutant-related risks in populations with particular characteristics (e.g., presence of pre-existing disease), or in particular species (i.e., for welfare effects), causality determinations reflect the strength of the evidence for effects in those populations or species.

5.3. Public Health Impacts

The public health impacts of criteria pollutant exposures can be influenced by the populations exposed, with higher-risk populations potentially experiencing more serious impacts, and by the severity of the effects in those populations. The ISA consideration of these issues is discussed below in sections 5.3.1 and 5.3.2, respectively.

5.3.1. At-Risk Populations and Lifestages

A critical part of assessing the public health impacts of criteria air pollutant exposures is the identification, evaluation, and characterization of populations and/or lifestages that may be at higher risk of pollutant-related effects (i.e., compared to the general population). Higher risks could be due to intrinsic, extrinsic, and/or exposure-related factors. Intrinsic factors such as genetics, lifestage, or disease status can contribute to larger or more severe responses to a particular pollutant exposure, and physiological differences between groups (e.g., differences in breathing patterns between children and adults) can contribute to higher internal pollutant doses in some populations. Extrinsic factors (e.g., nutritional status) and exposure-related factors (e.g., working outdoors, living near roadways or other pollution sources) can also contribute to increased risk, and many of the characteristics commonly used to classify populations potentially at greater risk (e.g., race, socioeconomic status, educational attainment) may be surrogates for the combined impact of several intrinsic, extrinsic, and exposure-related factors.

The ISA evaluation of factors that could increase risk draws from the evidence synthesis and integration underlying causality determinations (see section 5.1), with a focus on the subset of studies that provide information on pollutant-related effects in particular groups. This can include epidemiologic, controlled human exposure, and animal toxicological studies that provide information on pollutant-related effects in particular populations or lifestages, as well as studies examining differential exposures or dosimetry across populations or lifestages.

With regard to epidemiologic studies, ISAs focus on studies that include stratified analyses, which can compare effects across groups with similar air pollution exposures and within the same study design, and on studies that examine effects overwhelmingly or exclusively present in specific populations (e.g., lung function development in children; heart attacks or strokes in people with pre-existing cardiovascular disease). When evaluating results across epidemiologic studies, consistent with the approach to informing causality determinations, emphasis is placed on patterns or trends in results in the various populations evaluated (e.g., see section 5.1.1).

Careful consideration of how epidemiologic studies define groups is important to interpreting the evidence regarding potential at-risk populations or lifestages. For example, individuals with a pre-existing disease may exhibit different responses based on the extent to which the disease is being controlled. Population groups characterized as similar but defined by different metrics may also vary in risk (e.g., body mass index vs. other indicators of body composition, various indicators of socioeconomic status, and various age ranges used to define lifestages). A related consideration is variability in the response to pollutant exposures due to variation in behavior, biological characteristics, or adherence to medical treatments. The ISAs consider such sources of variation where relevant because they may affect the extent to which studies can reliably identify a population or lifestage at increased risk of pollutant-related effects, and thus the consistency of results across studies.

The ISAs also evaluate controlled human exposure and animal toxicological studies that examine potential risk factors, such as age or health status (e.g., pre-existing asthma). When available, these studies can provide information about the independent effects of exposure to the air pollutant under evaluation, as well as the biological plausibility of epidemiologic associations, in specific populations. A limitation of controlled human exposure studies is that they commonly exclude participants with serious health conditions, and thus generally do not reflect the highest-risk groups. Additionally, limitations in the generalizability of animal models to human health and disease often contribute important uncertainties to the findings of these studies.

The ISAs use information on factors that may place particular populations or lifestages at greater risk in several ways. Evidence for differential risk across groups could help explain heterogeneity in results across health studies that are conducted in different locations and/or populations. Additionally, evidence for effect measure modification could provide supporting information on mechanisms contributing to pollutant-related health effects, and consistent evidence that a particular population or lifestage is at greater risk of pollutant-related health effects can support causation in that population. For health outcomes with sufficient evidence, ISAs apply a structured framework to characterize potential risk factors and to describe the overall confidence that a specific factor may result in a population or lifestage being at higher risk of a

pollutant-related health effect. The ISA framework employs four categories to characterize the evidence for potential at-risk populations — "adequate evidence," "suggestive evidence," "inadequate evidence," and "evidence of no effect." These categories are described below in Table 9.

Table 9. Classification of evidence for factors potentially increasing the risk of criteria pollutant-related health effects.

Classification	Health Effects
Adequate evidence	There is substantial, consistent evidence within a discipline to conclude that a factor results in a population or lifestage being at higher risk of pollutant-related health effect(s) relative to some reference population or lifestage. Where applicable, this evidence includes coherence across disciplines. Evidence includes multiple high-quality studies.
Suggestive evidence	The collective evidence suggests that a factor results in a population or lifestage being at higher risk of pollutant-related health effect(s) relative to some reference population or lifestage, but the evidence is limited due to inconsistency within a discipline or, where applicable, a lack of coherence across disciplines.
Inadequate evidence	The collective evidence is inadequate to determine whether a factor results in a population or lifestage being at higher risk of pollutant-related health effect(s) relative to some reference population or lifestage. The available studies are of insufficient quantity, quality, consistency, and/or statistical power to permit a conclusion to be drawn.
Evidence of no effect	There is substantial, consistent evidence within a discipline to conclude that a factor does not result in a population or lifestage being at higher risk of pollutant-related health effect(s) relative to some reference population or lifestage. Where applicable, the evidence includes coherence across disciplines. Evidence includes multiple high-quality studies.

5.3.2. Severity of Effects

Many of the health outcomes evaluated in ISAs are clearly adverse (e.g., premature mortality, hospitalization, development or exacerbation of disease). However, some ISA outcomes reflect biological changes that are transient and reversible (e.g., temporary changes in lung function, blood pressure) or that provide mechanistic support for biological plausibility but are of uncertain clinical relevance by themselves (e.g., changes in biomarkers of oxidative stress or inflammation). For such responses, there may be no clearly identified threshold at which a biological change becomes severe enough to be of concern for health, and even small changes may indicate a

clinically significant response if they occur in combination with other effects or in less healthy populations. Interpretations of clinical significance can draw from factors such as established health impacts of particular response magnitudes, information on how response magnitudes and impacts may differ between the general population and groups at higher risk, an understanding of the natural variability in the response, and evidence supporting linkages to more overt manifestations of adversity. Where available, ISAs provide information to guide judgments about the potential adversity and public health implications of biological responses that are of uncertain clinical significance.[16]

For some clinical areas there may be professional scientific societies that have deliberated on the question of how to consider particular physiological or biomarker changes in the context of clinical significance and adversity. For example, the American Thoracic Society (ATS), an organization of respiratory disease specialists, has published several statements in this area (e.g., American Thoracic Society, 2000). The ATS described its 2000 statement as being intended to "provide guidance to policy makers and others who interpret the scientific evidence on the health effects of air pollution for the purposes of risk management" (ATS, 2000). A more recent joint statement of the ATS and the European Respiratory Society identifies "considerations to be weighed in setting boundaries between adverse and nonadverse health effects" (Thurston et al., 2017).

Broadly, these guidelines note that the clinical significance of a biological response can be influenced by the magnitude, duration, and frequency of the response and by the degree to which it may interfere with normal activity or result in illness, injury, or progressive dysfunction (Thurston et al., 2017). Additionally, the populations in which responses occur, and the extent to which pollutant exposures may shift the population risk distribution, can have implications for public health impacts (Thurston et al., 2017). Thus, when characterizing the potential adversity and public health implications of air pollutant exposures, important considerations include the magnitude and relative transience or persistence of responses, the impacts of responses on normal

[16] Judgments about adversity and public health impacts are often made in subsequent steps of the NAAQS review, including in the Policy Assessment and in proposed and final decisions.

activity and illness, and the pre-existing sensitivity of exposed populations (ATS, 2000; Thurston et al., 2017).

5.4. Public Welfare Impacts

Once a causality determination is made, answers to the following questions can inform the EPA's understanding of the broader public welfare impacts of criteria pollutant exposures:

- What ecosystem elements, endpoints, or services (e.g., types, regions, taxonomic groups, populations, and functions) appear to be affected or are more sensitive to effects?
- What exposure conditions (e.g., amount deposited or concentration, duration, and pattern) result in changes to welfare endpoints?
- What is the shape of the concentration-, exposure-, or dose-response relationship?
- Are there geographical differences in welfare effects responses?

In addressing these questions, ISAs evaluate the effects of criteria air pollutant exposures in sensitive species and ecosystems (section 5.4.1) and the severity of reported effects (section 5.4.2) in order to inform judgements on adversity and impacts on public welfare.

5.4.1. At-risk populations, species, ecosystems, and ecosystem services

A key aspect of evaluating the effects of criteria pollutant exposures on public welfare is identifying and assessing impacts on sensitive species, populations, and ecosystems (NASEM, 2022). Pollutant exposures can directly influence ecosystem endpoints, potentially resulting in critical impacts on at-risk species and habitats. Assessment of at-risk species and habitats includes consideration of both intrinsic and extrinsic factors. Intrinsic factors, such as physiological characteristics, genetics, life history traits, morphology, and species-specific tolerances, and extrinsic factors, like exposure conditions, environmental factors, and stressors, contribute to varying sensitivity and susceptibility of organisms and habitats to criteria air pollutants. For example, abundant species with physiological sensitivities, or species with limited ranges, low genetic diversity, or low population densities may be particularly vulnerable

to air pollution exposure. Similarly, ecosystems may be more at-risk when they have low biodiversity or are subjected to multiple, compounding environmental stressors including air pollutant exposure. Exposure assessment evaluates the intensity, duration, and frequency of contact between a stressor and the ecological receptors. This, combined with sensitivity analyses, defines vulnerability and determines whether a species or ecosystem is likely to be adversely affected. Sensitive ecological entities can include, but are not limited to, endangered, threatened, and rare species; National Park Service lands: Class I areas; federally designated critical habitats; and culturally important ecosystems. Assessing the impact of criteria air pollutants on sensitive species, populations, and ecosystems is important to understanding the public welfare impacts of criteria air pollutants (Kaylor et al., 2024), especially given the crucial services these ecosystems provide to the environment and people.

In addition to ecological endpoints, ecosystem services provide frameworks for evaluating how pollutant exposure may detrimentally impact species, habitats, or ecological functions of economic or cultural importance to humans. Ecosystem services are the benefits that people obtain directly or indirectly from ecosystems (Millennium Ecosystem Assessment (MEA), 2003), as human well-being is dependent on services provided by nature. In addressing the potential for effects on ecosystem services, ISAs recognize different types of ecosystem services including "provisioning services such as food and water; regulating services such as regulation of floods, drought, land degradation, and disease; supporting services such as soil formation and nutrient cycling; and cultural services such as recreational, spiritual, religious, and other nonmaterial benefits" (MEA, 2003). Several frameworks and tools have been formalized for the identification and valuation of ecosystem services and nature's contribution to people, which quantify the links between ecosystem endpoints and impacts on human well-being (e.g., MEA, 2003; TEEB, 2010, IPBES, 2019). The EPA's National Ecosystem Services Classification System-Plus (NECES Plus) tool assists users in tracing connections between ecosystem endpoints and human well-being by analyzing final ecosystem services (Newcomer-Johnson et al., 2020). The severity of criteria air pollutant impacts on public welfare depends on several factors, including extent of exposure, sensitivity of

ecological systems, type of ecosystem services affected, and the presence of modifying factors.

5.4.2. Severity of Effects

The CAA defines welfare effects broadly, encompassing diverse types of effects including environmental, ecological, economic, and social impacts. For example, CAA section 302(h) identifies impacts on vegetation as a public welfare effect. Effects on vegetation span multiple levels of biological organization, ranging from responses like changes in photosynthesis to large-scale shifts in ecosystem-level productivity. Synthesis of ecological endpoints across biological scales may provide additional insight into adversity (NASEM, 2022). Moreover, ecosystem-level shifts in vegetation have the potential to alter key ecological services, affecting air quality, biodiversity support, and recreational activities. Integrated Science Assessments inform judgments regarding the severity of effects on public welfare through the characterization of diverse effects across biological, spatial, and temporal scales.

Detrimental ecological effects of air pollutant exposures can occur across all levels of biological organization, spanning from suborganismal and organismal levels to population, community, and ecosystem levels. Although effects on an individual organism are unlikely to be considered adverse to public welfare, if effects occur on enough individuals within a population, then communities and ecosystems may be disrupted. Both short and long-term changes to populations, communities, and ecosystems can in turn result in an alteration of ecosystem functions. Population-level endpoints, such as growth, reproduction, and mortality rates, are directly linked to community and ecosystem health. Consequently, detrimental impacts of air pollutants on these population endpoints can negatively affect higher levels of biological organization. Other endpoints, such as changes in behavior and physiological stress, can decrease ecological fitness of an organism but are harder to link unequivocally to effects at the population, community, and ecosystem level.

Non-ecological welfare effects, such as impacts on visibility and materials damage, also affect ecosystem services and human and economic well-being. For example, adverse impacts on visibility may be expressed through aesthetic losses or economic impacts. Specifically, reduced visibility may impair aesthetic quality and

enjoyment of the visual environment, while also impacting property values. In addition, damage and soiling that impairs the function and aesthetic qualities of materials, including damage to property and cultural heritage sites, may be considered as detrimental impacts associated with pollutants in ambient air. Studies that characterize, and ideally quantify, how people perceive changes in visibility, the enjoyment of landscapes, and materials damage are important for effectively evaluating the impacts of criteria air pollutants on public welfare.

Effects of criteria air pollutants on Earth's radiative balance can lead to changes to the climate system, including near-surface air temperature, precipitation, and circulation patterns. These effects can be direct, such as cooling through the reflection of incoming solar radiation by sulfate aerosol, or indirect, such as changes to the properties of clouds by SO_2 or the role of CO in altering the concentrations of other greenhouse gases such as methane. While these impacts on the climate system also have downstream effects on terrestrial and aquatic environments and potential feedback loops, the direct impact of criteria pollutants on physical climate variables, such as radiative forcing or temperature, are the most evident metrics of detrimental impacts. Downstream effects on terrestrial ecosystems and any associated feedbacks, such as the adverse impacts of ozone on vegetation that impact CO_2 uptake and global temperature feedbacks, are more difficult to quantitatively attribute to changes in concentrations of individual pollutants regulated under the NAAQS.

6. PEER REVIEW AND PUBLIC COMMENT

Section 109(d)(2) of the CAA addresses the appointment and advisory functions of an independent scientific review committee. Among other requirements, section 109(d)(2) directs the Administrator to appoint this committee, which is to be composed of "seven members including at least one member of the National Academy of Sciences, one physician, and one person representing state air pollution control agencies." Section 109(d)(2) additionally provides that the independent scientific review committee "shall complete a review of the criteria...and the national primary and secondary ambient air quality standards…and shall recommend to the Administrator any new…standards and revisions of existing criteria and standards as may be appropriate…." Since the early

1980s, this independent review function has been performed by the CASAC of the EPA's Science Advisory Board.

In reviewing the draft ISA, the CASAC may be assisted by an ad hoc panel of independent subject matter experts. In past NAAQS reviews, ad hoc panels have been convened to provide broad expertise related to the particular criteria pollutant under evaluation and the science-policy issues important for the review. The CASAC convenes at public meetings that are announced in the *Federal Register* and that provide an opportunity for public comment. Draft advisory reports, conveying recommendations to the EPA on the draft ISA, are prepared by the CASAC or by the ad hoc panel for transmission to the chartered CASAC for discussion and deliberation. These reports convey advice to the EPA regarding the ISA's evidence assessment and policy-relevant conclusions. If the chartered CASAC determines the contents of a report are appropriate, the CASAC will adopt the report and transmit it to the EPA to reflect its statutorily mandated advice to the Agency.

The EPA carefully considers advice received from the CASAC and comments from the public on the draft ISA. This may include consideration of additional studies identified during CASAC's review that meet the ISA's scoping and study quality criteria. Final ISAs are made available on the EPA website, and a notice announcing the availability of the final ISA is published in the *Federal Register*. More information on the EPA's peer review practices can be found in EPA's Peer Review Handbook and via the EPA CASAC peer review website (https://casac.epa.gov/ords/sab/r/sab_apex/casac/home).

7. REFERENCES

American Thoracic Society (ATS). (2000). What constitutes an adverse health effect of air pollution? Am J Respir Crit Care Med 161: 665-673.

Bang, H and Robins, JM (2005). Doubly robust estimation in missing data and causal inference models. Biometrics 61(4): 962-972.

Barr, RG, Herbstman, J, Speizer, FE and Camargo, CA, Jr. (2002). Validation of self-reported chronic obstructive pulmonary disease in a cohort study of nurses. Am J Epidemiol 155(10): 965-971.

Bell ML, Dominici F, Samet JM. (2005). A meta-analysis of time-series studies of ozone and mortality with comparison to the National Morbidity, Mortality, and Air Pollution Study. Epidemiology. 16(4):436–445.

Chen W, Qian L, Shi J, Franklin M. Comparing performance between log-binomial and robust Poisson regression models for estimating risk ratios under model misspecification. BMC Med Res Methodol. 2018 Jun 22;18(1):63.

CDC (2004). National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention Atlanta (GA). Available at: https://pubmed.ncbi.nlm.nih.gov/24455788/.

Cronin, MTD, Bajot, F, Enoch, SJ, Madden, JC, Roberts, DW and Schwöbel, J (2009). The In Chemico–In Silico Interface: Challenges for Integrating Experimental and Computational Chemistry to Identify Toxicity. Alternatives to Laboratory Animals 37(5): 513-521.

Díaz, I (2019). Machine learning in the estimation of causal effects: Targeted minimum loss-based estimation and double/debiased machine learning. Biostatistics 21(2): 353-358.

Díaz, I, Williams, N, Hoffman, KL and Schenck, EJ (2023). Nonparametric causal effects based on longitudinal modified treatment policies. Journal of the American Statistical Association 118(542): 846-857.

ECHA (2016). New approach methodologies in regulatory science: Proceedings of a scientific workshop, Helsinki, 19–20 April 2016. Helsinki, Finland. Available at: http://dx.doi.org/10.2823/543644.

Ertefaie, A and Stephens, DA (2010). Comparing approaches to causal inference for longitudinal data: Inverse probability weighting versus propensity scores. The International Journal of Biostatistics 6(2): 14.

European Commission, Joint Research Centre, Adcock, I, Novotny, T, Nic, M, Dibusz, K, Hynes, J, Marshall, L and Gribaldo, L (2020). *Advanced non-animal models in biomedical research – Respiratory tract diseases*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2760/725821.

Gee, GC and Payne-Sturges, DC (2004). Environmental health disparities: A framework integrating psychosocial and environmental concepts. Environ Health Perspect 112(17): 1645-1653.

Gelman, A and Greenland, S (2019). Are confidence intervals better termed "uncertainty intervals"? BMJ (Clinical Research Edition) 366: l5381.

Greenland, S (2000). An introduction to instrumental variables for epidemiologists. Int J Epidemiol 29: 722-729.

Greenland, S, Senn, SJ, Rothman, KJ, Carlin, JB, Poole, C, Goodman, SN and Altman, DG (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31(4): 337-350.

Greenstone, M., & Gayer, T. (2009). Quasi-experimental and experimental approaches to environmental economics. J. Envir Econ Manag, 57(1), 21-44.

Hamel, C, Kelly, SE, Thavorn, K, Rice, DB, Wells, GA and Hutton, B (2020). An evaluation of DistillerSR's machine learning-based prioritization tool for title/abstract screening - impact on reviewer-relevant outcomes. BMC Med Res Methodology 20(1): 256.

Hernán, MA and Robins, JM (2016). Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 183(8): 758-764.

Hew (1964). Smoking and health: Report of the advisory committee to the surgeon general of the public health service. U.S. Department of Health, Education, and Welfare. Washington, DC. Available at: http://profiles.nlm.nih.gov/ps/retrieve/ResourceMetadata/NNBBMQ.

Hill, AB (1965). The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58(5): 295-300.

Howard, BE, Phillips, J, Tandon, A, Maharana, A, Elmore, R, Mav, D, Sedykh, A, Thayer, K, Merrick, BA, Walker, V, Rooney, A and Shah, RR (2020). SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. Environ Int 138: 105623.

IARC (2006). IARC monographs on the evaluation of carcinogenic risks to humans: Preamble. World Health Organization. Lyon, France.

IOM, Ed. (2008). Improving the presumptive disability decision-making process for veterans. National Academies Press Washington, DC.

IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages.

IPCC, 2014: Annex II: Glossary [Mach, K.J., S. Planton and C. von Stechow (eds.)]. In: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, pp. 117-130.

Kaylor, S.D., Dalton, R.M., Greaver, T. *et al.* (2024). Emerging Scientific Approaches for Identifying Ecologically Adverse Effects of Air Pollution. Environmental Management 74, 835–845.

Keil, AP, Buckley, JP, Katie, MO, Ferguson, KK, Zhao, S and White, AJ (2020). A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environ Health Perspect 128(4): 47004.

Lash, TL, Fox, MP, MacLehose, RF, Maldonado, G, McCandless, LC and Greenland, S (2014). Good practices for quantitative bias analysis. Int J Epidemiol 43(6): 1969-1985.

Millennium Ecosystem Assessment (MEA). (2003). Ecosystems and human well-being: a framework for assessment. Washington, DC: Island Press

Muhajarine, N, Mustard, C, Roos, LL, Young, TK and Gelskey, DE (1997). Comparison of survey and physician claims data for detecting hypertension. J Clin Epidemiol 50(6): 711-718.

Murgia, N, Brisman, J, Claesson, A, Muzi, G, Olin, AC and Torén, K (2014). Validity of a questionnaire-based diagnosis of chronic obstructive pulmonary disease in a general population-based study. BMC Pulmonary Medicine 14: 49.

NASEM (2017). *National Academies of Sciences, Engineering, and Medicine 2017. Controlled Human Inhalation-Exposure Studies at EPA*. National Academies Press (US). Washington (DC). Available at: http://dx.doi.org/10.17226/24618.

NASEM (2022). National Academies of Sciences, Engineering, and Medicine 2022. Advancing the Framework for Assessing Causality of Health and Welfare Effects to Inform National Ambient Air Quality Standard Reviews. The National Academies Press. Washington, DC. Available at: https://doi.org/10.17226/26612.

Newcomer-Johnson, T., F. Andrews, J. Corona, Ted DeWitt, M. Harwell, C. Rhodes, P. Ringold, M. Russell, P. Sinha, AND G. Van Houtven. (2020). National Ecosystem Services Classification System (NESCS Plus). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-20/267, 2020.

Pearce, Neil; Vandenbroucke, Jan P.; Lawlor, Deborah A. (2019). Causal Inference in Environmental Epidemiology: Old and New Approaches. Epidemiology 30(3):p 311-316.

Richardson, A, Hudgens, MG, Gilbert, PB and Fine, JP (2014). Nonparametric bounds and sensitivity analysis of treatment effects. Statistical Science 29(4): 596-618.

Richmond-Bryant and Long. Influence of exposure measurement errors on results from epidemiologic studies of different designs. Journal of Exposure Science & Environmental Epidemiology, 2020, 30(3), 420-429.

Robins, JM, Hernán, MA and Brumback, B (2000). Marginal structural models and causal inference in epidemiology. Epidemiology 11(5): 550-560.

Rothman, KJ and Greenland, S (1998). *Modern Epidemiology*. Lippincott-Raven. Philadelphia, PA.

Savitz, DA and Forastiere, F (2021). Do pooled estimates from meta-analyses of observational epidemiology studies contribute to causal inference? Occup Environ Med 78(9): 621-622.

Savitz, DA, Wellenius, GA and Trikalinos, TA (2019). The Problem with Mechanistic Risk of Bias Assessments in Evidence Synthesis of Observational Studies and a Practical Alternative: Assessing the Impact of Specific Sources of Potential Bias. Am J Epidemiol 188(9): 1581-1585.

Sheppard, L, Slaughter, JC, Schildcrout, J, Liu, JS and Lumley, T (2005). Exposure and measurement contributions to estimates of acute air pollution effects. J Expo Anal Environ Epidemiol 15(4): 366-376.

Stitelman, OM, De Gruttola, V and van der Laan, MJ (2012). A general implementation of TMLE for longitudinal data applied to causal inference in survival analysis. The International Journal of Biostatistics 8(1): 26.

Stuart, EA (2010). Matching methods for causal inference: A review and a look forward. Statistical Science 25(1): 1-21.

Talbot D, Mésidor M, Chiu Y, Simard M, Sirois C. An Alternative Perspective on the Robust Poisson Method for Estimating Risk or Prevalence Ratios. Epidemiology. 2023 Jan 1;34(1):1-7.

TEEB (2010) The Economics of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: A synthesis of the approach, conclusions and recommendations of TEEB.

Thurston GD, Kipen H, Annesi-Maesano I, et al. (2017). A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. Eur Respir J; 49: 1600419.

Toren, K, Brisman, J and Jarvholm, B (1993). Asthma and asthma-like symptoms in adults assessed by questionnaires: A literature review. Chest 104(2): 600-608.

U.S. EPA (1986). Air quality criteria for lead: Volume II of IV [EPA Report]. Research Triangle Park, NC. EPA-600/8-83/028aF Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=32647.

U.S. EPA (1991). Guidelines for Developmental Toxicity Risk Assessment. U.S. Environmental Protection Agency. Available at: https://nepis.epa.gov/exe/ZyPURL.cgi?Dockey=30002L7L.txt.

U.S. EPA (1996). Guidelines for Reproductive Toxicity Risk Assessment. U.S. Environmental Protection Agency, Risk Assessment Forum. Washington, DC. EPA/630/R-96/009. Available at: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=30004YQB.txt.

U.S. EPA (1998). Guidelines for neurotoxicity risk assessment. U.S. Environmental Protection Agency, Risk Assessment Forum. Washington, DC. EPA/630/R-95/001F. Available at: http://www.epa.gov/risk/guidelines-neurotoxicity-risk-assessment.

U.S. EPA (2005). Guidelines for carcinogen risk assessment. Risk Assessment Forum. Washington, DC. EPA630P03001F. Available at: https://www.epa.gov/sites/production/files/201309/documents/cancer_guidelines_final_3-25-05.pdf.

U.S. EPA (2015). Preamble to the Integrated Science Assessments (ISA). Office of Research and Development, National Center for Environmental Assessment. Washington, DC. U.S. EPA. EPA/600/R-15/067. Available at: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=526136.

U.S. EPA (2016). Integrated Science Assessment for Oxides of Nitrogen-Health Criteria. U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. Research Triangle Park, NC. EPA/600/R-15/068. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310879.

U.S. EPA (2018). Strategic plan to promote the development and implementation of alternative test methods within the TSCA program. Office of Chemical Safety and Pollution Prevention. Washington D.C. EPA-740-R1-8004. Available at: https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf.

U.S. EPA (2019a). Integrated Science Assessment (ISA) for Particulate Matter (Final Report). U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. Washington, DC. U.S. EPA. EPA/600/R-19/188. December 2019. Available at: https://www.epa.gov/naaqs/particulate-matter-pm-standards-integrated-science-assessments-current-review.

U.S. EPA (2019b). Guidelines for Human Exposure Assessment. Risk Assessment Forum. Washington, DC. EPA/100/B-19/001. Available at: https://www.epa.gov/sites/production/files/2020-01/documents/guidelines_for_human_exposure_assessment_final2019.pdf.

U.S. EPA (2020a). Integrated Science Assessment for Oxides of Nitrogen, Oxides of Sulfur, and Particulate Matter—Ecological Criteria (Final). U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment. Research Triangle Park, NC. EPA/600/R-20/278. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=349473.

U.S. EPA (2020b). Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final). U.S. Environmental Protection Agency, Office of

Research and Development. Washington, DC. EPA/600/R-20/012. Available at: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=540022.

U.S. EPA (2021). EPA new approach methods work plan: reducing use of vertebrate animals in chemical testing. Washington D.C. Available at: https://www.epa.gov/chemical-research/epa-new-approach-methods-work-plan-reducing-use-vertebrate-animals-chemical.

U.S. EPA (2022). ORD staff handbook for developing IRIS assessments. U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment. Washington, DC. EPA 600/R-22/268. Available at: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=356370.

U.S. EPA (2023). Living Literature Review [Computer Software], RTP, NC. https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/11327944

U.S. EPA (2024a). Integrated Science Assessment (ISA) for Lead (Final Report). Office of Research and Development, Center for Public Health and Environmental Assessment. Washington, DC. U.S. EPA. EPA/600/R-23/375. January 2024. Available at: https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=548355.

U.S. EPA (2024b). Integrated Review Plan for the Primary National Ambient Air Quality Standards for Oxides of Nitrogen. Volume 2: Planning for the Review and the Integrated Science Assessment. Office of Research and Development and Office of Air and Radiation. Washington, DC. U.S. EPA. EPA-452/R-24-010b. March 2024. Available at: https://www.epa.gov/system/files/documents/2024-03/no2-irp-volume-2_march-2024.pdf.

U.S. EPA (2024c). Integrated Review Plan for the National Ambient Air Quality Standards for Ozone and Related Photochemical Oxidants Volume 2: Planning for the Review and the Integrated Science Assessment. Office of Research and Development and Office of Air and Radiation. Washington, DC. U.S. EPA. EPA-452/R-24-001b. December 2024. Available at: https://www.epa.gov/system/files/documents/2024-12/o3-irp-vol-2app_final.pdf.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med, 167(4), 268-274.

Vinikoor-Imler, L. C., Owens, E. O., Nichols, J. L., Ross, M., Brown, J. S., & Sacks, J. D. (2014). Evaluating potential response-modifying factors for associations between

ozone and health outcomes: a weight-of-evidence approach. Environ Health Perspect, 122(11), 1166-1176.

Vinikoor-Imler, LC, Owens, EO, Nichols, JL, Ross, M, Brown, JS and Sacks, JD (2014). Evaluating potential response-modifying factors for associations between ozone and health outcomes: A weight-of-evidence approach. Environ Health Perspect, 122(11): 1166-1176.

Wasserstein, RL and Lazar, NA (2016). The ASA's statement on p-values: Context, process, and purpose. American Statistician 70(2): 129-133.

Wasserstein, RL, Schirm, AL and Lazar, NA (2019). Moving to a world beyond "p<0.05". American Statistician 73(sup1): 1-19.

Weakley, J, Webber, MP, Ye, F, Zeig-Owens, R, Cohen, HW, Hall, CB, Kelly, K and Prezant, DJ (2013). Agreement between obstructive airways disease diagnoses from self-report questionnaires and medical records. Preventive Medicine 57(1): 38-42.

Wei, Y; Qiu, X; Yazdi, MD; Shtein, A; Shi, L; Yang, J; Peralta, AA; Coull, BA; Schwartz, JD. (2022). The Impact of Exposure Measurement Error on the Estimated Concentration-Response Relationship between Long-Term Exposure to $PM_{2.5}$ and Mortality. Environ Health Perspect 130: 77006. http://dx.doi.org/10.1289/EHP10389.

Weuve, J, Sagiv, SK and Fox, MP (2018). Quantitative bias analysis for collaborative science. Epidemiology 29(5): 627-630.

WHO (2008). Uncertainty and Data Quality in Exposure Assessment. The International Programme on Chemical Safety. Geneva. WHO. Available at: https://apps.who.int/iris/handle/10665/44017.

WHO (2021). *World Health Organization global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization. Geneva. Available at: https://apps.who.int/iris/handle/10665/345329.

Wing, C, Simon, K and Bello-Gomez, RA (2018). Designing difference in difference studies: Best practices for public health policy research. Annu Rev Public Health 39(1): 453-469.

Yoo EH, Pu Q, Eum Y, Jiang X. The Impact of Individual Mobility on Long-Term Exposure to Ambient $PM_{2.5}$: Assessing Effect Modification by Travel Patterns and Spatial Variability of $PM_{2.5}$. Int J Environ Res Public Health. 2021 Feb 23;18(4):2194.

Zeger, SL, Thomas, D, Dominici, F, Samet, JM, Schwartz, J, Dockery, D and Cohen, A (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. Environ Health Perspect 108(5): 419-426.

Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702-6.

Zou GY, Donner A. Extension of the modified Poisson regression model to prospective studies with correlated binary data. Stat Methods Med Res. 2013 Dec;22(6):661-70.