PREAMBLE

Process of ISA Development

This preamble outlines the general process for developing an Integrated Science Assessment (ISA) including the framework for evaluating weight of evidence and drawing scientific conclusions and causal judgments. The ISA provides a concise review, synthesis, and evaluation of the most policy-relevant science to serve as a scientific foundation for the review of the National Ambient Air Quality Standards (NAAQS). The general process for NAAQS reviews is described at http://www.epa.gov/ttn/naaqs/review.html. Figure I depicts the general NAAQS review process and information for individual NAAQS reviews is available at www.epa.gov/ttn/naaqs. This preamble is a general discussion of the basic steps and criteria used in developing an ISA; for each ISA, specific details and considerations are included in the introductory section for that assessment.

The fundamental process for developing an ISA includes:

- literature searches;
- study selection;
- evaluation and integration of the evidence;
- development of scientific conclusions and causal judgments.

An initial step in this process is publication of a call for information in the Federal Register that invites the public to provide information relevant to the assessment, such as new or recent publications on health or welfare[1] effects of the pollutant, or from atmospheric and exposure sciences fields. EPA maintains an ongoing literature search process for identification of relevant scientific studies published since the last review of the NAAQS. Search strategies are designed for pollutants and scientific disciplines and iteratively modified to optimize identification of pertinent publications. Papers are identified for inclusion in several additional ways: specialized searches on specific topics; independent review of tables of contents for journals in which relevant papers may be published; independent identification of relevant literature by expert scientists; review of citations in previous assessments and identification by the public and the Clean Air Scientific Advisory Committee (CASAC) during the external review process. This literature search and study selection process is depicted in Figure II. Publications considered for inclusion in the ISA are added to the Health and Environmental Research Online (HERO) database developed by EPA (http://hero.epa.gov/); the references in the ISA include a hyperlink to the database.

---

[1] Welfare effects as defined in Clean Air Act (CAA) section 302(h) [42 U.S.C. 7602(h)] include, but are not limited to, “effects on soils, water, crops, vegetation, man-made materials, animals, wildlife, weather, visibility and climate, damage to and deterioration of property, and hazards to transportation, as well as effects on economic values and on personal comfort and well-being.”

Studies that have undergone scientific peer review and have been published or accepted for publication and reports that have undergone review are considered for inclusion in the ISA. Analyses conducted by EPA using publicly available data are also considered for inclusion in the ISA. All relevant epidemiologic, controlled human exposure, toxicological, and ecological and welfare effects studies published since the last review are considered, including those related to exposure-response relationships, mode(s) of action (MOA), and potentially at-risk populations and lifestages. Studies on atmospheric chemistry, environmental fate and transport, dosimetry, toxicokinetics and exposure are also considered for inclusion in the document, as well as analyses of air quality and emissions data. References that were considered for inclusion in a specific ISA can be found using the HERO website (http://hero.epa.gov).

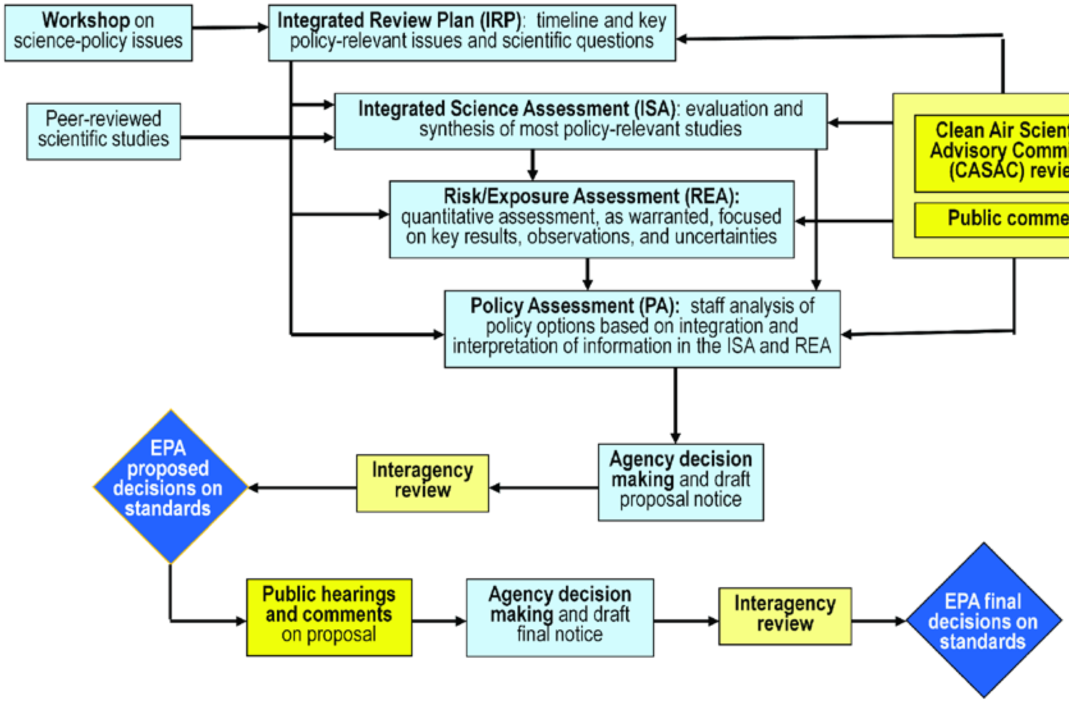
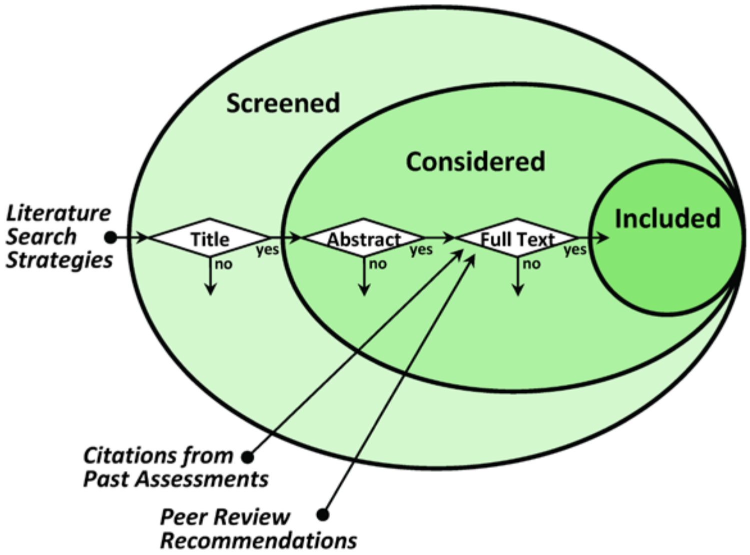


Figure I Illustration of the key steps in the process of the review of National Ambient Air Quality Standards.



Criteria for study evaluation include:

- Are the study populations, subjects, or animal models adequately selected, and are they sufficiently well defined to allow for meaningful comparisons between study or exposure groups?
- Are the statistical analyses appropriate, properly performed, and properly interpreted? Are likely covariates adequately controlled or taken into account in the study design and statistical analysis?
- Are the air quality data, exposure, or dose metrics of adequate quality and sufficiently representative of information regarding ambient conditions?
- Are the health, ecological or welfare effect measurements meaningful, valid and reliable?
- Do the analytical methods provide adequate sensitivity and precision to support conclusions?

Figure II Illustration of processes for literature search and study selection used for development of ISAs.

Each ISA builds upon the conclusions of previous assessments for the pollutant under review. EPA focuses on peer reviewed literature published following the completion of the previous review (2006 O_3 AQCD) and on any new interpretations of previous literature, integrating the results of recent scientific studies with previous findings. Important earlier studies may be discussed in detail to reinforce key concepts and conclusions or for reinterpretation in light of newer data. Earlier studies also are the primary focus in some areas of the document where research efforts have subsided, or if these earlier studies remain the definitive works available in the literature.

Selection of studies for inclusion in the ISA is based on the general scientific quality of the study, and consideration of the extent to which the study is informative and policy-relevant. Policy relevant and informative studies include those that provide a basis for or describe the relationship between the criteria pollutant and effects, including studies that offer innovation in method or design and studies that reduce uncertainty on critical issues, such as analyses of confounding or effect modification by copollutants or other variables, analyses of concentration-response or dose-response relationships, or analyses related to time between exposure and response. Emphasis is placed on studies that examine effects associated with pollutant concentrations relevant to current population and ecosystem exposures, and particularly those pertaining to concentrations currently found in ambient air. Other studies are included if they contain unique data, such as a previously unreported effect or MOA for an observed effect, or examine multiple concentrations to elucidate exposure-response relationships. In general, in assessing the scientific quality and relevance of health and welfare effects studies, the following considerations have been taken into account when selecting studies for inclusion in the ISA.

- Are the study populations, subjects, or animal models adequately selected, and are they sufficiently well defined to allow for meaningful comparisons between study or exposure groups?
- Are the statistical analyses appropriate, properly performed, and properly interpreted? Are likely covariates adequately controlled or taken into account in the study design and statistical analysis?
- Are the air quality data, exposure, or dose metrics of adequate quality and sufficiently representative of information regarding ambient conditions?
- Are the health, ecological or welfare effect measurements meaningful, valid and reliable?
- Do the analytical methods provide adequate sensitivity and precision to support conclusions?

Considerations specific to particular disciplines include the following: In selecting epidemiologic studies, EPA considers whether a given study: (1) presents information on associations with short- or long-term pollutant exposures at or near conditions relevant to ambient exposures; (2) addresses potential confounding by other pollutants; (3) assesses potential effect modifiers; (4) evaluates health endpoints and populations not previously extensively researched; and (5) evaluates important methodological issues related to interpretation of the health evidence (e.g., lag or time period between exposure and effects, model specifications, thresholds, mortality displacement).

Considerations for the selection of research evaluating controlled human exposure or animal toxicological studies include a focus on studies conducted using relevant pollutant exposures. For both types of studies, relevant pollutant exposures are considered to be those generally within one or two orders of magnitude of ambient concentrations. Studies in which higher doses were used may also be considered if they provide information relevant to understanding MOA or mechanisms, as noted below.

Evaluation of controlled human exposure studies focuses on those that approximated expected human exposure conditions in terms of concentration and duration. Studies should include control exposures to filtered air, as appropriate. In the selection of controlled human exposure studies, emphasis is placed on studies that: (1) investigate potentially at-risk populations and lifestages such as people with asthma or cardiovascular diseases, children or older adults; (2) address issues such as concentration-response or time-course of responses; and (3) have sufficient statistical power to assess findings.

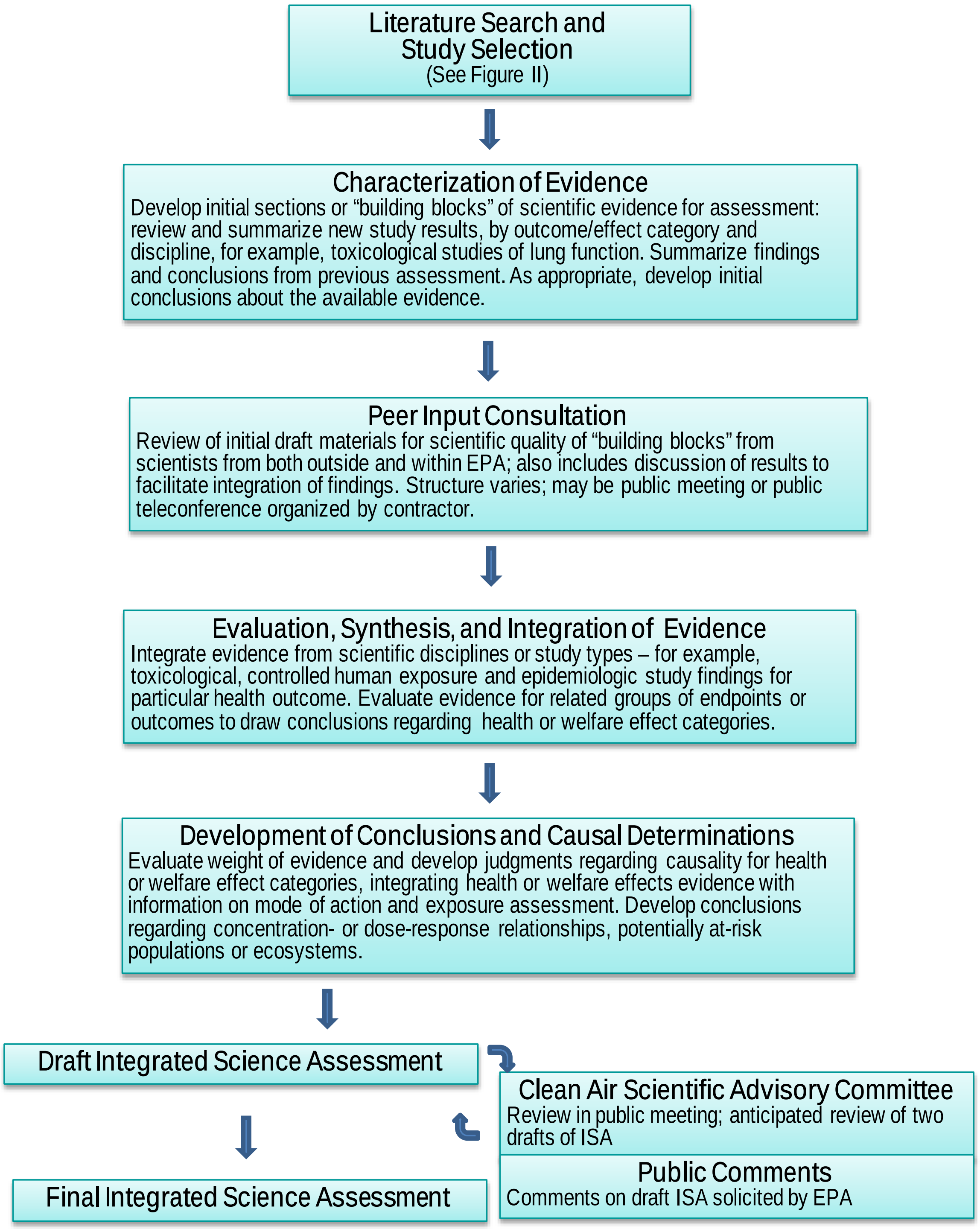
Review of the animal toxicological evidence focuses on studies that approximate expected human dose conditions, which vary depending on the dosimetry, toxicokinetics, and biological sensitivity of the particular laboratory animal species or strains studied. Emphasis is placed on studies that: (1) investigate animal models of disease that can provide information on populations potentially at increased risk of effects; (2) address issues such as concentration-response or time-course of responses; and (3) have sufficient statistical power to assess findings. Due to resource constraints on exposure duration and numbers of animals tested, animal studies typically utilize high-concentration exposures to acquire data relating to mechanisms and assure a measurable response. Emphasis is placed on studies using doses or concentrations generally within 1-2 orders of magnitude of current levels. Studies with higher concentration exposures or doses are considered to the extent that they provide useful information to inform understanding of interspecies differences between healthy and at-risk human populations. Results from in vitro studies may also be included if they provide mechanistic insight or further support for results demonstrated in vivo.

These criteria provide benchmarks for evaluating various studies and for focusing on the policy-relevant studies in assessing the body of health, ecological and welfare effects evidence. As stated initially, the intent of the ISA is to provide a concise review, synthesis, and evaluation of the most policy-relevant science to serve as a scientific foundation for the review of the NAAQS, not extensive summaries of all health, ecological and welfare effects studies for a pollutant. Of most relevance for inclusion of studies is whether they provide useful qualitative or quantitative information on exposure-effect or exposure-response relationships for effects associated with pollutant exposures at doses or concentrations relevant to ambient conditions that can inform decisions on whether to retain or revise the standards.

The general process for ISA development is illustrated in Figure III. In developing an ISA, EPA reviews and summarizes the evidence from studies of atmospheric sciences; human exposure, toxicological, controlled human exposure and epidemiologic studies; and studies of ecological and welfare effects. In the process of developing the first draft ISA, EPA may convene a peer input meeting in which EPA the scientific content of preliminary draft materials is reviewed to ensure that the ISA is up to date and focused on the most policy-relevant findings, and to assist EPA with integration of evidence within and across disciplines. EPA integrates the evidence from across scientific disciplines or study types and characterizes the weight of evidence for relationships between the pollutant and various outcomes.

The integration of evidence on health, and ecological or welfare effects, involves collaboration between scientists from various disciplines. As an example, an evaluation of health effects evidence would include the integration of the results from epidemiologic, controlled human exposure, and toxicological studies, and application of the causal framework (described below) to draw conclusions. Integration of results on health or ecological effects that are logically or mechanistically connected (e.g., a spectrum of effects on the respiratory system) informs judgments of causality. Using the causal framework described in the following section, EPA scientists consider aspects such as strength, consistency, coherence, and biological plausibility of the evidence, and develop causality determinations on the nature of the relationships. Causality determinations often entail an iterative process of review and evaluation of the evidence. Two drafts of the ISA are typically released for review by the CASAC and the public, and comments received on the characterization of the science as well as the implementation of the causal framework are carefully considered in revising and completing the final ISA.

Integrated Science Assessment Development Process



Literature Search and Study Selection
(See Figure II)

Characterization of Evidence
Develop initial sections or “building blocks” of scientific evidence for assessment: review and summarize new study results, by outcome/effect category and discipline, for example, toxicological studies of lung function. Summarize findings and conclusions from previous assessment. As appropriate, develop initial conclusions about the available evidence.

Peer Input Consultation
Review of initial draft materials for scientific quality of “building blocks” from scientists from both outside and within EPA; also includes discussion of results to facilitate integration of findings. Structure varies; may be public meeting or public teleconference organized by contractor.

Evaluation, Synthesis, and Integration of Evidence
Integrate evidence from scientific disciplines or study types – for example, toxicological, controlled human exposure and epidemiologic study findings for particular health outcome. Evaluate evidence for related groups of endpoints or outcomes to draw conclusions regarding health or welfare effect categories.

Development of Conclusions and Causal Determinations
Evaluate weight of evidence and develop judgments regarding causality for health or welfare effect categories, integrating health or welfare effects evidence with information on mode of action and exposure assessment. Develop conclusions regarding concentration- or dose-response relationships, potentially at-risk populations or ecosystems.

Draft Integrated Science Assessment

Clean Air Scientific Advisory Committee
Review in public meeting; anticipated review of two drafts of ISA

Public Comments
Comments on draft ISA solicited by EPA

Final Integrated Science Assessment

Figure III Characterization of the general process of ISA development.

EPA Framework for Causal Determination

EPA has developed a consistent and transparent basis for integration of scientific evidence and evaluation of the causal nature of air pollution-related health or welfare effects for use in developing ISAs. The framework described below establishes uniform language concerning causality and brings more specificity to the findings. This standardized language was drawn from sources across the federal government and wider scientific community, especially the National Academy of Sciences (NAS) Institute of Medicine (IOM) document, *Improving the Presumptive Disability Decision-Making Process for Veterans* (Samet and Bodurow, 2008), a comprehensive report on evaluating causality. This framework:

- describes the kinds of scientific evidence used in establishing a general causal relationship between exposure and health effects;
- characterizes the process for integration and evaluation of evidence necessary to reach a conclusion about the existence of a causal relationship;
- identifies issues and approaches related to uncertainty;
- provides a framework for classifying and characterizing the weight of evidence in support of a general causal relationship.

Approaches to assessing the separate and combined lines of evidence (e.g., epidemiologic, controlled human exposure, and animal toxicological studies) have been formulated by a number of regulatory and science agencies, including the IOM of the NAS (Samet and Bodurow, 2008), International Agency for Research on Cancer (IARC, 2006), U.S. EPA (2005), and Centers for Disease Control and Prevention (CDC, 2004). Causal inference criteria have also been described for ecological effects evidence (U.S. EPA, 1998; Fox, 1991). These formalized approaches offer guidance for assessing causality. The frameworks are similar in nature, although adapted to different purposes, and have proven effective in providing a uniform structure and language for causal determinations.

Evaluating Evidence for Inferring Causation

The 1964 Surgeon General’s (U.S. Department of Health, Education and Welfare [HEW]) 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 defined as an agent that brings about an effect or a result. An association is the statistical relationship among variables; alone, however, it is insufficient proof of a causal relationship between an exposure and a health outcome. 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 (Samet and Bodurow, 2008).

Many of the health and environmental outcomes reported in these studies have complex etiologies. Diseases such as asthma, coronary heart disease (CHD) or cancer are typically initiated by multiple agents. Outcomes depend on a variety of factors, such as age, genetic susceptibility, nutritional status, immune competence, and social factors (Samet and Bodurow, 2008; Gee and Payne-Sturges, 2004). Effects on ecosystems are often also multifactorial with a complex web of causation. Further, exposure to a combination of agents could cause synergistic or antagonistic effects. Thus, the observed risk may represent the net effect of many actions and counteractions.

Scientific findings incorporate uncertainty. “Uncertainty” can be defined as having limited knowledge to exactly describe an existing state or future outcome, e.g., the lack of knowledge about the correct value for a specific measure or estimate. Uncertainty analysis may be qualitative or quantitative in nature. In many cases, the analysis is qualitative, and can include professional judgment or inferences based on analogy with similar situations. Quantitative uncertainty analysis may include use of simple measures (e.g., ranges) and analytical techniques. Quantitative uncertainty analysis might progress to more complex measures and techniques, if needed for decision support. Various approaches to evaluating uncertainty include classical statistical methods, sensitivity analysis, or probabilistic uncertainty analysis, in order of increasing complexity and data requirements. However, data may not be available for all aspects of an assessment and those data that are available may be of questionable or unknown quality. Ultimately, the assessment is based on a number of assumptions with varying degrees of uncertainty.

Publication bias is a source of uncertainty regarding the magnitude of health risk estimates. It is well understood that studies reporting non-null findings are more likely to be published than reports of null findings. Publication bias can result in overestimation of effect estimate sizes (Ioannidis, 2008). For example, effect estimates from single-city epidemiologic studies have been found to be generally larger than those from multicity studies which is an indication of publication bias in that null or negative single-city results may be reported in a multicity analyses but might not be published independently (Bell et al., 2005).

Consideration of Evidence from Scientific Disciplines

Moving from association to causation involves the elimination of alternative explanations for the association. The ISA focuses on evaluation of the findings from the body of evidence, drawing upon the results of all studies determined to meet the criteria described previously. Causality determinations are based on the evaluation, integration, and synthesis of evidence from across scientific disciplines. The relative importance of different types of evidence varies by pollutant or assessment, as does the availability of different types of evidence for causality determination. Three general types of studies inform consideration of human health effects: controlled human exposure, epidemiologic, and toxicological studies. Evidence on ecological or welfare effects may be drawn from a variety of experimental approaches (e.g., greenhouse, laboratory, field) and numerous disciplines (e.g., community ecology, biogeochemistry, and paleontological/historical reconstructions).

Direct evidence of a relationship between pollutant exposures and human health effects comes from controlled human exposure studies. Such studies experimentally evaluate the health effects of administered exposures in human volunteers under highly controlled laboratory conditions. Also referred to as human clinical studies, these experiments allow investigators to expose subjects to known concentrations of air pollutants under carefully regulated environmental conditions and activity levels. These studies provide important information on the biological plausibility of associations observed in epidemiologic studies. Essential dose-response profiles and ranges of response severity can be established with these studies. In some instances, controlled human exposure studies can also be used to characterize concentration-response relationships at pollutant concentrations relevant to ambient conditions. Controlled human exposures are typically conducted using a randomized crossover design, with subjects exposed both to the pollutant and a clean air control. In this way, subjects serve as their own controls, effectively controlling for many potential confounders. Considerations for evaluating controlled human study findings include the generally small sample size and short exposure time used in experimental studies, and that severe health outcomes are not assessed. By experimental design, controlled human exposure studies are structured to evaluate physiological or biomolecular outcomes in response to exposure to a specific air pollutant and/or combination of pollutants. In addition, the study design generally precludes inclusion of subjects with serious health conditions, and therefore the results often cannot be generalized to an entire population. Although some controlled human exposure studies have included health-compromised individuals such as those with respiratory or cardiovascular disease, these individuals may also be relatively healthy and may not represent the most sensitive individuals in the population. Thus, observed effects in these studies may underestimate the response in certain populations.

Epidemiologic studies provide important information on the associations between health effects and exposure of human populations to ambient air pollution.
In epidemiologic or observational studies of humans, the investigator generally does not control exposures or intervene with the study population. Broadly, observational studies can describe associations between exposures and effects. These studies fall into several categories: e.g., cross-sectional, prospective cohort, panel, and time-series studies. Cross-sectional studies use health outcome, exposure and covariate data available at the community level (e.g., annual mortality rates and pollutant concentrations), but do not have individual-level data. Prospective cohort studies have some data collected at the individual level, generally health outcome data, and in some cases individual-level data on exposure and covariates are collected. Time-series studies evaluate the relationship for changes in a health outcome with changes in exposure indicators, such as an association between daily changes in mortality with air pollution. Panel studies include repeated measurements of health outcomes, such as respiratory symptoms or heart rhythm variable, at the individual level. “Natural experiments” offer the opportunity to investigate changes in health related to a change in exposure, such as closure of a pollution source.

In evaluating epidemiologic studies, consideration of many study design factors and issues must be taken into account to properly inform their interpretation. One key consideration is evaluation of the potential contribution of the pollutant to a health outcome when it is a component of a complex air pollutant mixture. Reported effect estimates in epidemiologic studies may reflect: independent effects on health outcomes; effects of the pollutant acting as an indicator of a copollutant or a complex ambient air pollution mixture; effects resulting from interactions between that pollutant and copollutants.

In the evaluation of epidemiologic evidence, one important consideration is 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). One approach to remove spurious associations (due to possible confounders); is to control for characteristics that may differ between exposed and unexposed persons; this is frequently termed "adjustment." Scientific judgment is needed to evaluate likely sources and extent of confounding, together with consideration of how well the existing constellation of study designs, results, and analyses address the potential for erroneous inferences. A confounder is associated with both the exposure and the effect; for example, confounding can occur between correlated pollutants that are associated with the same effect.

Several statistical methods are available to detect and control for potential confounders, with none of them being completely satisfactory. Multivariable regression models constitute one tool for estimating the association between exposure and outcome after adjusting for characteristics of participants that might confound the results. The use of multipollutant regression models has been the prevailing approach for controlling potential confounding by copollutants in air pollution health effects studies. Finding the likely causal pollutant from multipollutant regression models is made difficult by the possibility that one or more air pollutants may be acting as a surrogate for an unmeasured or poorly measured pollutant or for a particular mixture of pollutants. In addition, pollutants may independently exert effects on the same system; for example, several pollutants may be associated with respiratory effects through either the same or different modes of action. The number and degree of diversity of covariates, as well as their relevance to the potential confounders, remain matters of scientific judgment. Despite these limitations, the use of multipollutant models is still the prevailing approach employed in most air pollution epidemiologic studies and provides some insight into the potential for confounding or interaction among pollutants.

Confidence that unmeasured confounders are not producing the findings is increased when multiple studies are conducted in various settings using different subjects or exposures, each of which might eliminate another source of confounding from consideration. For example, multicity studies can provide insight on potential confounding through the use of a consistent method to analyze data from across locations with different levels of copollutants and other covariates. Intervention studies, because of their quasi-experimental nature, can be particularly useful in characterizing causation.

Another important consideration in the evaluation of epidemiologic evidence is effect modification, which occurs when the effect differs between subgroups or strata; for example, effect estimates that vary by age group or potential risk factor. As stated by Rothman and Greenland (1998):

> "Effect-measure modification differs from confounding in several ways. The main difference is that, whereas confounding is a bias that the investigator hopes to prevent or remove from the effect estimate, effect-measure modification is a property of the effect under study …
> In epidemiologic analysis one tries to eliminate confounding but one tries to detect and estimate effect-measure modification."

When a risk factor is a confounder, it is the true cause of the association observed between the exposure and the outcome; when a risk factor is an effect modifier, it changes the magnitude of the association between the exposure and the outcome in stratified analyses. For example, the presence of a pre-existing disease or indicator of low socioeconomic status may act as effect modifiers if they are associated with increased risk of effects related to air pollution exposure. It is often possible to stratify the relationship between health outcome and exposure by one or more of these potential effect modifiers. For variables that modify the association, effect estimates in each stratum will be different from one another and different from the overall estimate, indicating a different exposure-response relationship may exist in populations represented by these variables.

Exposure measurement error, which refers to the uncertainty associated with the exposure metrics used to represent exposure of an individual or population, can be an important contributor to uncertainty in air pollution epidemiologic study results. Exposure error can influence observed epidemiologic associations between ambient pollutant concentrations and health outcomes by biasing effect estimates toward or away from the null and widening confidence intervals around those estimates (Zeger et al., 2000). There are several components that contribute to exposure measurement error in air pollution epidemiologic studies, including the difference between true and measured ambient concentrations, the difference between average personal exposure to ambient pollutants and ambient concentrations at central monitoring sites, and the use of average population exposure rather than individual exposure estimates. Factors that could influence exposure estimates include nonambient sources of exposure, topography of the natural and built environment, meteorology, measurement errors, time-location-activity patterns, and the extent to which ambient pollutants penetrate indoor environments. The importance of exposure error varies with study design and is dependent on the spatial and temporal aspects of the design.

The third main type of health effects evidence, animal toxicological studies, provides information on the pollutant's biological action under controlled and monitored exposure circumstances. Taking into account physiological differences of the experimental species from humans, these studies inform characterization of health effects of concern, exposure-response relationships and MOAs. Further, animal models can inform determinations of at-risk populations. These studies evaluate the effects of exposures to a variety of pollutants in a highly controlled laboratory setting and allow exploration of toxicological pathways or mechanisms by which a pollutant may cause effects. Understanding the biological mechanisms underlying various health outcomes can prove crucial in establishing or negating causality. In the absence of human studies data, extensive, well-conducted animal toxicological studies can support determinations of causality, if the evidence base indicates that similar responses are expected in humans under ambient exposure conditions.

Interpretations of animal toxicological studies are affected by limitations associated with extrapolation between animal and human responses. The differences between humans and other species have to be taken into consideration, including metabolism, hormonal regulation, breathing pattern, and differences in lung structure and anatomy. Also, in spite of a high degree of homology and the existence of a high percentage of orthologous genes across humans and rodents (particularly mice), extrapolation of molecular alterations at the gene level is complicated by species-specific differences in transcriptional regulation. Given these differences, there are uncertainties associated with quantitative extrapolations of observed pollutant-induced pathophysiological alterations between laboratory animals and humans, as those alterations are under the control of widely varying biochemical, endocrine, and neuronal factors.

For ecological effects assessment, both laboratory and field studies (including field experiments and observational studies) can provide useful data for causality determination. Because conditions can be controlled in laboratory studies, responses may be less variable and smaller differences may be easier to detect. However, the control conditions may limit the range of responses (e.g., animals may not be able to seek alternative food sources) or incompletely reflect pollutant bioavailability, so they may not reflect responses that would occur in the natural environment. In addition, larger-scale processes are difficult to reproduce in the laboratory.

Field observational studies measure biological changes in uncontrolled situations, and describe an association between a disturbance and an ecological effect. Field data can provide important information for assessments of multiple stressors or where site-specific factors significantly influence exposure. They are also often useful for analyses of larger geographic scales and higher levels of biological organization. However, because conditions are not controlled, variability is expected to be higher and differences harder to detect. Field surveys are most useful for linking stressors with effects when stressor and effect levels are measured concurrently. The presence of confounding factors can make it difficult to attribute observed effects to specific stressors.

Intermediate between laboratory and field are studies that use environmental media collected from the field to examine response in the laboratory, and experiments that are performed in the natural environment while controlling for some environmental conditions (i.e., mesocosm studies). This type of study in manipulated natural environments can be considered a hybrid between a field experiment and laboratory study since some aspects are performed under controlled conditions but others are not. They make it possible to observe community and/or ecosystem dynamics, and provide strong evidence for causality when combined with findings of studies that have been made under more controlled conditions.

Application of Framework for Causal Determination

In its evaluation and integration of the scientific evidence on health or welfare effects of criteria pollutants, EPA determines the weight of evidence in support of causation and characterizes the strength of any resulting causal classification. EPA also evaluates the quantitative evidence and draws scientific conclusions, to the extent possible, regarding the concentration-response relationships and the loads to ecosystems, exposures, doses or concentrations, exposure duration, and pattern of exposures at which effects are observed.

To aid judgment, various "aspects"[1] of causality have been discussed by many philosophers and scientists. 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 aspects of causality in epidemiology and public health that have been widely used (Samet and Bodurow, 2008; IARC, 2006; U.S. EPA, 2005; CDC, 2004). These aspects (Hill, 1965) have been modified (Table I) for use in causal determinations specific to health and welfare effects for pollutant exposures (U.S. EPA, 2009d).[2] Although these aspects provide a framework for assessing the evidence, they do not lend themselves to being considered in terms of simple formulas or fixed rules of evidence leading to conclusions about causality (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 the evidence and the likelihood of causality. Rather, these aspects provide a framework for systematic appraisal of the body of evidence, informed by peer and public comment and advice, which includes weighing alternative views on controversial issues. In addition, it is important to note that the aspects in Table I cannot be used as a strict checklist, but rather to determine the weight of the evidence for inferring causality. In particular, not meeting one or more of the principles does not automatically preclude a determination of causality [see discussion in CDC (2004)].

---

[1] The "aspects" described by Sir Austin Bradford Hill (Hill, 1965) have become, in the subsequent literature, more commonly described as "criteria." The original term "aspects" is used here to avoid confusion with "criteria" as it is used, with different meaning, in the Clean Air Act.

[2] The Hill aspects were developed for interpretation of epidemiologic results. They have been modified here for use with a broader array of data, i.e., epidemiologic, controlled human exposure, ecological, and animal toxicological studies, as well as in vitro data, and to be more consistent with the U.S. EPA (2005) Guidelines for Carcinogen Risk Assessment.

Table I Aspects to aid in judging causality.

Aspect	Description
Consistency of the observed association	An inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. The reproducibility of findings constitutes one of the strongest arguments for causality. 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 causality from one line of evidence (e.g., epidemiologic, controlled human exposure [clinical], or animal studies) may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association. Evidence on ecological or welfare effects may be drawn from a variety of experimental approaches (e.g., greenhouse, laboratory, and field) and subdisciplines of ecology (e.g., community ecology, biogeochemistry, and paleontological/historical reconstructions). The coherence of evidence from various fields greatly adds to the strength of an inference of causality. In addition, there may be coherence in demonstrating effects across multiple study designs or related health endpoints within one scientific line of evidence.
Biological plausibility	An inference of causality tends to be strengthened by consistency with data from experimental studies or other sources demonstrating plausible biological mechanisms. A proposed mechanistic linking between an effect and exposure to the agent is an important source of support for causality, especially when data establishing the existence and functioning of those mechanistic links are available.
Biological gradient (exposure-response relationship)	A well-characterized exposure-response relationship (e.g., increasing effects associated with greater exposure) strongly suggests cause and effect, especially when such relationships are also observed for duration of exposure (e.g., increasing effects observed following longer exposure times).
Strength of the observed association	The finding of large, precise risks increases confidence that the association is not likely due to chance, bias, or other factors. However, it is noted that a small magnitude in an effect estimate may represent a substantial effect in a population.
Experimental evidence	Strong evidence for causality can be provided through "natural experiments" when a change in exposure is found to result in a change in occurrence or frequency of health or welfare effects.
Temporal relationship of the observed association	Evidence of a temporal sequence between the introduction of an agent, and appearance of the effect, constitutes another argument in favor of causality.
Specificity of the observed association	Evidence linking a specific outcome to an exposure can provide a strong argument for causation. However, it must be recognized that rarely, if ever, does exposure to a pollutant invariably predict the occurrence of an outcome, and that a given outcome may have multiple causes.
Analogy	Structure activity relationships and information on the agent's structural analogs can provide insight into whether an association is causal. Similarly, information on mode of action for a chemical, as one of many structural analogs, can inform decisions regarding likely causality.

Determination of Causality

In the ISA, EPA assesses the body of relevant literature, building upon evidence available during previous NAAQS reviews, to draw conclusions on the causal relationships between relevant pollutant exposures and health or environmental effects. ISAs use a five-level hierarchy that classifies the weight of evidence for causation[1]. In developing this hierarchy, EPA has drawn on the work of previous evaluations, most prominently the IOM's *Improving the Presumptive Disability Decision-Making Process for Veterans* (Samet and Bodurow, 2008), EPA's Guidelines for Carcinogen Risk Assessment (U.S. EPA, 2005), and the U.S. Surgeon General's smoking report (CDC, 2004). This weight of evidence evaluation is based on integration of findings from various lines of evidence from across the health and environmental effects disciplines. These separate judgments are integrated into a qualitative statement about the overall weight of the evidence and causality. The five descriptors for causal determination are described in Table II.

Determination of causality involves the evaluation and integration of evidence for different types of health, ecological or welfare effects associated with short- and long-term exposure periods. In making determinations of causality, evidence is evaluated for major outcome categories or groups of related endpoints (e.g., respiratory effects, vegetation growth), integrating evidence from across disciplines, and assessing the coherence of evidence across a spectrum of related endpoints to draw conclusions regarding causality. In discussing the causal determination, EPA characterizes the evidence on which the judgment is based, including strength of evidence for individual endpoints within the outcome category or group of related endpoints.

In drawing judgments regarding causality for the criteria air pollutants, the ISA focuses on evidence of effects in the range of relevant pollutant exposures or doses, and not on determination of causality at any dose. Emphasis is placed on evidence of effects at doses (e.g., blood Pb concentration) or exposures (e.g., air concentrations) that are relevant to, or somewhat above, those currently experienced by the population. The extent to which studies of higher concentrations are considered varies by pollutant and major outcome category, but generally includes those with doses or exposures in the range of one to two orders of magnitude above current or ambient conditions. Studies that use higher doses or exposures may also be considered to the extent that they provide useful information to inform understanding of mode of action, interspecies differences, or factors that may increase risk of effects for a population. Thus, a causality determination is based on weight of evidence evaluation for health, ecological or welfare effects, focusing on the evidence from exposures or doses generally ranging from current levels to one or two orders of magnitude above current levels.

---

[1] Both the CDC and IOM frameworks use a four-category hierarchy for the strength of the evidence. A five-level hierarchy is used here to be consistent with the EPA Guidelines for Carcinogen Risk Assessment and to provide a more nuanced set of categories.

In addition, EPA evaluates evidence relevant to understand the quantitative relationships between pollutant exposures and health, ecological or welfare effects. This includes evaluation of the form of concentration-response or dose-response relationships and, to the extent possible, drawing conclusions on the levels at which effects are observed. The ISA also draws scientific conclusions regarding important exposure conditions for effects and populations that may be at greater risk for effects, as described in the following section.

Table II Weight of evidence for causal determination.

	Health Effects	Ecological and Welfare Effects
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (i.e., doses or exposures generally within one to two orders of magnitude of current levels). That is, the pollutant has been shown to result in health effects in studies in which chance, bias, and confounding could be ruled out with reasonable confidence. For example: a) controlled human exposure studies that demonstrate consistent effects; or b) observational studies that cannot be explained by plausible alternatives or are supported by other lines of evidence (e.g., animal studies or mode of action information). Evidence includes multiple high-quality studies	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures i.e., doses or exposures generally within one to two orders of magnitude of current levels). That is, the pollutant has been shown to result in effects in studies in which chance, bias, and confounding could be ruled out with reasonable confidence. Controlled exposure studies (laboratory or small- to medium-scale field studies) provide the strongest evidence for causality, but the scope of inference may be limited. Generally, determination is based on multiple studies conducted by multiple 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.
Likely to be a causal relationship	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures, but important uncertainties remain. That is, the pollutant has been shown to result in health effects in studies in which chance and bias can be ruled out with reasonable confidence but potential issues remain. For example: a) observational studies show an association, but copollutant exposures are difficult to address and/or other lines of evidence (controlled human exposure, animal, or mode of action information) are limited or inconsistent; or b) animal toxicological evidence from multiple studies from different laboratories that demonstrate effects, but limited or no human data are available. Evidence generally includes multiple high-quality studies.	Evidence is sufficient to conclude that there is a likely causal association with relevant pollutant exposures. That is, an association has been observed between the pollutant and the outcome in studies in which chance, bias, and confounding are minimized, but uncertainties remain. For example, field studies show a relationship, but suspected interacting factors cannot be controlled, and other lines of evidence are limited or inconsistent. Generally, determination is based on multiple studies in multiple research groups.
Suggestive of a causal relationship	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but is limited. For example, (a) at least one high-quality epidemiologic study shows an association with a given health outcome but the results of other studies are inconsistent; or (b) a well-conducted toxicological study, such as those conducted in the National Toxicology Program (NTP), shows effects in animal species,	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but chance, bias and confounding cannot be ruled out. For example, at least one high-quality study shows an effect, but the results of other studies are inconsistent.
Inadequate to infer a causal relationship	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quantity, quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.	The available studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.
Not likely to be a causal relationship	Evidence is suggestive of no causal relationship with relevant pollutant exposures. Several adequate studies, covering the full range of levels of exposure that human beings are known to encounter and considering at-risk populations, are mutually consistent in not showing an effect at any level of exposure.	Several adequate studies, examining relationships with relevant exposures, are consistent in failing to show an effect at any level of exposure.

Quantitative Relationships: Effects on Human Populations

Once a determination is made regarding the causal relationship between the pollutant and outcome category, important questions regarding quantitative relationships include:

- What is the concentration-response, exposure-response, or dose-response relationship in the human population?
- What is the interrelationship between incidence and severity of effect?
- What exposure conditions (dose or exposure, duration and pattern) are important?
- What populations and lifestages appear to be differentially affected (i.e., more at risk of experiencing effects)?

To address these questions, the entirety of quantitative evidence is evaluated to characterize pollutant concentrations and exposure durations at which effects were observed for exposed populations, including populations and lifestages potentially at increased risk. To accomplish this, evidence is considered from multiple and diverse types of studies, and a study or set of studies that best approximates the concentration-response relationships between health outcomes and the pollutant may be identified. Controlled human exposure studies provide the most direct and quantifiable exposure-response data on the human health effects of pollutant exposures. To the extent available, the ISA evaluates results from across epidemiologic studies that characterize the form of relationships between the pollutant and health outcomes and draws conclusions on the shape of these relationships. Animal data may also inform evaluation of concentration-response relationships, particularly relative to MOAs and characteristics of at-risk populations.

An important consideration in characterizing the public health impacts associated with exposure to a pollutant is whether the concentration-response relationship is linear across the range of concentrations or if nonlinear relationships exist along any part of this range. Of particular interest is the shape of the concentration-response curve at and below the level of the current standards. Various sources of variability and uncertainty, such as low data density in the lower concentration range, possible influence of exposure measurement error, and variability between individuals in susceptibility to air pollution health effects, tend to smooth and “linearize” the concentration-response function, and thus can obscure the existence of a threshold or nonlinear relationship [2006 O_3 AQCD (U.S. EPA, 2006b)]. Since individual thresholds vary from person to person due to individual differences such as genetic level susceptibility or pre-existing disease conditions (and even can vary from one time to another for a given person), it can be difficult to demonstrate that a threshold exists in a population study. These sources of variability and uncertainty may explain why the available human data at ambient concentrations for some environmental pollutants (e.g., particulate matter [PM], O_3, lead [Pb], environmental tobacco smoke [ETS], radiation) do not exhibit thresholds for cancer or noncancer health effects, even though likely mechanisms include nonlinear processes for some key events.

Finally, identification of the population groups or lifestages that may be at greater risk of health effects from air pollutant exposures contributes to an understanding of the public health impact of pollutant exposures. In the ISA, the term “at-risk population” is used to encompass populations or lifestages that have a greater likelihood of experiencing health effects related to exposure to an air pollutant due to a variety of factors; other terms used in the literature include susceptible, vulnerable, and sensitive. These factors may be intrinsic, such as genetic or developmental factors, race, sex, lifestage, or the presence of pre-existing diseases, or they may be extrinsic, such as socioeconomic status (SES), activity pattern and exercise level, reduced access to health care, low educational attainment, or increased pollutant exposures (e.g., near roadways). Epidemiologic studies can help identify populations potentially at increased risk of effects by evaluating health responses in the study population. Examples include testing for interactions or effect modification by factors such as sex, age group, or health status. Experimental studies using animal models of susceptibility or disease can also inform the extent to which health risks are likely greater in specific population groups.

Quantitative Relationships: Effects on Ecosystems or Public Welfare

Key questions for understanding the quantitative relationships between exposure (or concentration or deposition) to a pollutant and risk to ecosystems or the public welfare include:

- What elements of the ecosystem (e.g., types, regions, taxonomic groups, populations, functions, etc.) appear to be affected, or are more sensitive to effects? Are there differences between locations or materials in welfare effects responses, such as impaired visibility or materials damage?
- Under what exposure conditions (amount deposited or concentration, duration and pattern) are effects seen?
- What is the shape of the concentration-response or exposure-response relationship?

Evaluations of causality generally consider the probability of quantitative changes in ecological and welfare effects in response to exposure. A challenge to the quantification of exposure-response relationships for ecological effects is the great regional and local spatial variability, as well as temporal variability, in ecosystems. Thus, exposure-response relationships are often determined for a specific ecological system and scale, rather than at the national or even regional scale. Quantitative relationships therefore are estimated site by site and may differ greatly between ecosystems.

Concepts in Evaluating Adversity of Health Effects

In evaluating health evidence, a number of factors can be considered in delineating between adverse and nonadverse health effects resulting from exposure to air pollution. Some health outcomes, such as hospitalization for respiratory or cardiovascular diseases, are clearly considered adverse. It is more difficult to determine the extent of change that constitutes adversity in more subtle health measures. These include a wide variety of responses, such as alterations in markers of inflammation or oxidative stress, changes in pulmonary function or heart rate variability, or alterations in neurocognitive function measures. The challenge is determining the magnitude of change in these measures when there is no clear point at which a change becomes adverse. The extent to which a change in health measure constitutes an adverse health effect may vary between populations. Some changes that may not be considered adverse in healthy individuals would be potentially adverse in more at-risk individuals.

The extent to which changes in lung function are adverse has been discussed by the American Thoracic Society (ATS) in an official statement titled *What Constitutes an Adverse Health Effect of Air Pollution?* (ATS, 2000b). An air pollution-induced shift in the population distribution of a given risk factor for a health outcome was viewed as adverse, even though it may not increase the risk of any one individual to an unacceptable level. For example, a population of asthmatics could have a distribution of lung function such that no identifiable individual has a level associated with significant impairment. Exposure to air pollution could shift the distribution such that no identifiable individual experiences any clinically relevant effects. This shift toward decreased lung function, however, would be considered adverse because individuals within the population would have diminished reserve function and therefore would be at increased risk to further environmental insult. The committee also observed that elevations of biomarkers, such as cell number and types, cytokines and reactive oxygen species, may signal risk for ongoing injury and clinical effects or may simply indicate transient responses that can provide insights into mechanisms of injury, thus illustrating the lack of clear boundaries that separate adverse from nonadverse effects.

The more subtle health outcomes may be connected mechanistically to health events that are clearly adverse. For example, air pollution may affect markers of transient myocardial ischemia such as ST-segment abnormalities and onset of exertional angina. These effects may not be apparent to the individual, yet may still increase the risk of a number of cardiac events, including myocardial infarction and sudden death. Thus, small changes in physiological measures may not appear to be clearly adverse when considered alone, but may be a part of a coherent and biologically plausible chain of related health outcomes that range up to responses that are very clearly adverse, such as hospitalization or mortality.

Concepts in Evaluating Adversity of Ecological Effects

Adversity of ecological effects can be understood in terms ranging in biological level of organization; from the cellular level to the individual organism and to the population, community, and ecosystem levels. In the context of ecology, a population is a group of individuals of the same species, and a community is an assemblage of populations of different species interacting with one another that inhabit an area.

An ecosystem is the interactive system formed from all living organisms and their abiotic (physical and chemical) environment within a given area (IPCC, 2007a). The boundaries of what could be called an ecosystem are somewhat arbitrary, depending on the focus of interest or study. Thus, the extent of an ecosystem may range from very small spatial scales to, ultimately, the entire Earth (IPCC, 2007a).

Effects on an individual organism are generally not considered to be adverse to public welfare. However if effects occur to enough individuals within a population, then communities and ecosystems may be disrupted. Changes to populations, communities, and ecosystems can in turn result in an alteration of ecosystem processes. Ecosystem processes are defined as the metabolic functions of ecosystems including energy flow, elemental cycling, and the production, consumption and decomposition of organic matter (U.S. EPA, 2002). Growth, reproduction, and mortality are species-level endpoints that can be clearly linked to community and ecosystem effects and are considered to be adverse when negatively affected. 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. The degree to which pollutant exposure is considered adverse may also depend on the location and its intended use (i.e., city park, commercial, cropland). Support for consideration of adversity beyond the species level by making explicit the linkages between stress-related effects at the species and effects at the ecosystem level is found in A *Framework for Assessing and Reporting on Ecological Condition: an SAB report* (U.S. EPA, 2002). Additionally, the National Acid Precipitation Assessment Program (NAPAP, 1991) uses the following working definition of “adverse ecological effects” in the preparation of reports to Congress mandated by the Clean Air Act: “any injury (i.e., loss of chemical or physical quality or viability) to any ecological or ecosystem component, up to and including at the regional level, over both long and short terms.”

On a broader scale, ecosystem services may provide indicators for ecological impacts. Ecosystem services are the benefits that people obtain from ecosystems (UNEP, 2003). According to the Millennium Ecosystem Assessment, ecosystem services include: “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.” For example, a more subtle ecological effect of pollution exposure may result in a clearly adverse impact on ecosystem services if it results in a population decline in a species that is recreationally or culturally important.

References

ATS (American Thoracic Society). (2000b). What constitutes an adverse health effect of air pollution? This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. Am J Respir Crit Care Med 161: 665-673.

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: 436-445. http://dx.doi.org/10.1097/01.ede.0000165817.40152.85

CDC (Centers for Disease Control and Prevention). (2004). The health consequences of smoking: A report of the Surgeon General. Washington, DC: U.S. Department of Health and Human Services. http://www.surgeongeneral.gov/library/smokingconsequences/

Fox, GA. (1991). Practical causal inference for ecoepidemiologists. J Toxicol Environ Health A 33: 359-373. http://dx.doi.org/10.1080/15287399109531535

Gee, GC; Payne-Sturges, DC. (2004). Environmental health disparities: A framework integrating psychosocial and environmental concepts. Environ Health Perspect 112: 1645-1653. http://dx.doi.org/10.1289/ehp.7074

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

Hill, AB. (1965). The environment and disease: Association or causation? Proc R Soc Med 58: 295-300.

IARC (International Agency for Research on Cancer). (2006). Preamble to the IARC monographs. Lyon, France. http://monographs.iarc.fr/ENG/Preamble/

Ioannidis, JPA. (2008). Why most discovered true associations are inflated [Review]. Epidemiology 19: 640-648. http://dx.doi.org/10.1097/EDE.0b013e31818131e7

IPCC (Intergovernmental Panel on Climate Change). (2007a). Climate change 2007: Impacts, adaptation and vulnerability. Cambridge, UK: Cambridge University Press.

NAPAP (National Acid Precipitation Assessment Program). (1991). The experience and legacy of NAPAP: Report of the Oversight Review Board of the National Acid Precipitation Assessment Program. Washington, DC.

Rothman, KJ; Greenland, S. (1998). Modern epidemiology (2nd ed.). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Samet, JM; Bodurow, CC. (2008). Improving the presumptive disability decision-making process for veterans. In JM Samet; CC Bodurow (Eds.). Washington, DC: National Academies Press. http://www.nap.edu/openbook.php?record_id=11908

U.S. EPA (U.S. Environmental Protection Agency). (1998). Guidelines for ecological risk assessment [EPA Report]. (EPA/630/R-95/002F). Washington, DC. http://www.epa.gov/raf/publications/guidelines-ecological-risk-assessment.htm

U.S. EPA (U.S. Environmental Protection Agency). (2002). A framework for assessing and reporting on ecological condition: An SAB report [EPA Report]. (EPA-SAB-EPEC-02-009). Washington, DC. http://www.ntis.gov/search/product.aspx?ABBR=PB2004100741

U.S. EPA (U.S. Environmental Protection Agency). (2005). Guidelines for carcinogen risk assessment [EPA Report]. (EPA/630/P-03/001F). Washington, DC. http://www.epa.gov/cancerguidelines/

U.S. EPA (U.S. Environmental Protection Agency). (2006b). Air quality criteria for ozone and related photochemical oxidants [EPA Report]. (EPA/600/R-05/004AF). Research Triangle Park, NC. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=149923

U.S. EPA (U.S. Environmental Protection Agency). (2009d). Integrated science assessment for particulate matter [EPA Report]. (EPA/600/R-08/139F). Research Triangle Park, NC. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=216546

UNEP (United Nations Environment Programme). (2003). Ecosystems and human well-being: A framework for assessment. Washington, DC: Island Press.

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