Published in final edited form as:
Regul Toxicol Pharmacol. 2017 August ; 88: 332–337. doi:10.1016/j.yrtph.2017.05.014.

FRAMEWORK FOR ASSESSING CAUSALITY OF AIR POLLUTION-RELATED HEALTH EFFECTS FOR REVIEWS OF THE NATIONAL AMBIENT AIR QUALITY STANDARDS

Elizabeth Oesterling Owens[a], **Molini M. Patel**[a], **Ellen Kirrane**[a], **Thomas C. Long**[a], **James Brown**[a], **Ila Cote**[a], **Mary A. Ross**[a], and **Steven J. Dutton**[a]

[a]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

Abstract

To inform regulatory decisions on the risk due to exposure to ambient air pollution, consistent and transparent communication of the scientific evidence is essential. The United States Environmental Protection Agency (U.S. EPA) develops the Integrated Science Assessment (ISA), which contains evaluations of the policy-relevant science on the effects of criteria air pollutants and conveys critical science judgments to inform decisions on the National Ambient Air Quality Standards. This article discusses the approach and causal framework used in the ISAs to evaluate and integrate various lines of scientific evidence and draw conclusions about the causal nature of air pollution-induced health effects. The framework has been applied to diverse pollutants and cancer and noncancer effects. To demonstrate its flexibility, we provide examples of causality judgments on relationships between health effects and pollutant exposures, drawing from recent ISAs for ozone, lead, carbon monoxide, and oxides of nitrogen. U.S. EPA's causal framework has increased transparency by establishing a structured process for evaluating and integrating various lines of evidence and uniform approach for determining causality. The framework brings consistency and specificity to the conclusions in the ISA, and the flexibility of the framework makes it relevant for evaluations of evidence across media and health effects.

Keywords

criteria air pollutants; causal determination; hazard identification; national ambient air quality standards

Refereeing and Publication Corresponding Author Elizabeth Oesterling Owens, National Center for Environmental Assessment, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive (MS A-110), Cincinnati, OH 45268, Phone: 513-569-7241; owens.beth@epa.gov. **Post-publication Corresponding Author** Steven J. Dutton, National Center for Environmental Assessment, U.S. Environmental Protection Agency, MD B243-01, Research Triangle Park, NC 27711, Phone: 919.541.5035; dutton.steven@epa.gov.

Competing interests
The authors declare they have no competing financial interests.

1. INTRODUCTION

The Clean Air Act requires the United States Environmental Protection Agency (U.S. EPA) to establish National Ambient Air Quality Standards (NAAQS) and to review the NAAQS on a periodic basis (CAA 1990a, 1990b). There are NAAQS for six "criteria" air pollutants: carbon monoxide (CO); lead (Pb); oxides of nitrogen; ozone (O_3); particulate matter; and sulfur oxides. The Integrated Science Assessment (ISA) provides the scientific foundation for the NAAQS, containing evaluations and integration of policy-relevant scientific evidence and communicating critical science judgments about the relationships between a criteria pollutant and health effects.

For the ISA, U.S. EPA developed a framework to support the consistent and transparent evaluation of evidence and determination of causality (i.e., extent to which cause-effect relationship can be inferred from a body of evidence) for criteria pollutant-induced health (cancer and noncancer) and welfare effects. This framework was first applied in the 2008 Integrated Science Assessment for Oxides of Nitrogen – Health Criteria (U.S. EPA 2008) and has evolved through application in ISAs for each of the criteria pollutants and with input from the public and the Clean Air Scientific Advisory Committee (CASAC)—an independent panel of scientific experts charged with reviewing the ISA. The framework establishes a uniform approach, enhancing consistency and specificity of causal determinations and allowing for transparent discourse between individuals when differences in the interpretation of scientific evidence arise. Here, we characterize the process for evaluating human health evidence and the framework used in the ISA for causal determinations [The Preamble to the ISA describes the full process, including that for evaluating welfare effects (U.S. EPA 2015)]. We also provide examples to illustrate how the framework is implemented and to show its emphasis on evidence integration across scientific disciplines and flexibility of application across diverse evidence types and health effects.

1.1. The ISA Process

1.1.1. Study Identification and Selection—U.S. EPA conducts a systematic literature search to identify recent peer-reviewed studies and reports not considered in the previous review for each criteria pollutant. Peer-reviewed studies and reports are the focus of ISA evaluations so that causal determinations are based on the best available science. Search strategies are designed *a priori* and iteratively modified to optimize identification of relevant publications. CASAC and the public may also recommend publications. A key factor in selecting a study for inclusion in the ISA is whether it provides useful qualitative or quantitative information on exposure-response relationships for effects associated with pollutant exposures at doses or concentrations relevant to ambient conditions. The ISA also includes U.S. EPA analyses of publicly available pollutant emissions and air quality data for characterizing the sources and ambient concentrations of a pollutant.

1.1.2. Evaluation of Individual Study Quality—After relevant studies are selected for inclusion, study quality is evaluated considering design, methods, and documentation, but not results. Specific aspects can include exposure and outcome assessment, consideration

of confounding and alternative explanations, and statistical methodology. This approach aims to consider the strengths and limitations, as well as the roles of chance, confounding, and other biases (e.g., selection bias, measurement error) that may affect interpretation of the study. Study quality assessment informs selection of studies to emphasize for evidence integration and causal determination, but the presence or absence of particular features does not necessarily define a study's utility or exclude it from consideration. A study that is judged to have limitations or be lower in quality would not generally be excluded from consideration in the ISA.

1.1.3. Evaluation and Integration of Evidence across Disciplines—Key to determining causality of relationships between criteria pollutant exposure and health effects is U.S. EPA's approach for integrating evidence from epidemiologic, controlled human exposure, and animal toxicological studies, including mode-of-action information. Evidence for similar and related health effects is evaluated and integrated across disciplines and between the current and prior NAAQS reviews. This process emphasizes consideration of consistency in the pattern of effects as well as strengths and limitations in the overall evidence. The roles and availability of different types of evidence in drawing causal determinations can vary by health effect and pollutant. Evaluation and integration of evidence also includes consideration of uncertainty, which is inherent in scientific findings. Uncertainty analysis may be qualitative or quantitative in nature. Generally, uncertainties and sources of heterogeneity in the evidence base are assessed qualitatively in the ISA, as understanding of these issues can be well-informed by integrating information across study designs, health outcomes, and disciplines. Where available, published meta-analyses are included in this qualitative analyses, and for some ISAs, U.S. EPA has conducted its own quantitative analysis approaches such as meta-regression. Publication bias as a potential source of uncertainty is considered during the evaluation and integration of evidence.

Integration of evidence across disciplines can strengthen causal inferences. Each study design and discipline has its strengths and limitations. A weak inference from one line of evidence can be offset by other lines of evidence, and coherence of these lines of evidence can add support to a cause-effect interpretation of an association. Interpretation of the body of scientific information as evidence of causal relationships involves assessing the full evidence base to rule out alternative explanations for observed effects with reasonable confidence.

1.1.4. Framework for Causal Determinations—In evaluating and integrating evidence on health effects of criteria pollutants, U.S. EPA considers various "aspects" of causality, principally those proposed by Bradford Hill (Hill 1965). These aspects, which include consistency, coherence, and biological plausibility (Table 1), have been modified by the U.S. EPA for use in the ISAs to be applicable to a broader array of data (i.e., epidemiologic, controlled human exposure, animal toxicological). These aspects provide a basis for systematically appraising the breadth of evidence but do not lead to simple formulas or fixed rules for forming causal conclusions (Hill 1965). Not meeting one or more of the aspects does not necessarily preclude a determination of a causal relationship (CDC 2004). Rather, these aspects are considered when making judgments on the body of evidence

as a whole and deciding what causal determination in the causal framework hierarchy best matches those judgments. The causal determinations and judgments of the evidence, which includes weighing alternative views on controversial issues, are informed by peer and public comment and advice.

Based on the characterizations of the body of evidence, U.S. EPA determines causality according to a five-level hierarchy that is consistent with EPA Guidelines for Carcinogen Risk Assessment (U.S. EPA 2005). Table 2 and the examples in the discussion describe the weight of evidence for each level. For a "*causal relationship*," chance, confounding and other biases can be ruled out with reasonable confidence. Lower in the hierarchy, from "*likely to be a causal relationship*" to "*inadequate to infer a causal relationship*", there is increasing uncertainty in the evidence linking pollutant exposure to a health effect. Sources of uncertainty include inconsistent findings, limited coherence across scientific disciplines, and/or alternative explanations that cannot be ruled out. For "*not likely to be a causal relationship*," multiple studies consistently show no effect. For any given causal determination, the evidence base is not required to comprise studies of a defined level of quality. For example, the determination of a "*causal relationship*" does not require all studies to be high quality. In making judgments on the body of evidence, studies of higher quality (i.e., with less uncertainty) are emphasized and studies with greater uncertainty would not be a major influence in determining causality if higher quality studies are available. However, studies with limitations may still provide useful information to address uncertainties on a specific issue. As the evidence base for a given health effect increasingly comprises studies with large uncertainty, the determination of causality will be lower in the hierarchy.

Causality is determined for broad health categories (e.g., respiratory effects) or groups of related effects (e.g., cognitive function). The final determinations are based on consensus, formed initially by ISA authors representing expertise across disciplines and developed over drafts of the ISA with consideration of public comment and CASAC review, which provide advice on the appropriate interpretation of the evidence. U.S. EPA focuses on the evidence of effects examined in relation to pollutant exposures or doses considered relevant to ambient conditions (generally within two orders of magnitude of current concentrations) and does not determine causality at any particular dose. In discussing the rationale for a causal determination, U.S. EPA characterizes the evidence on which the judgment is based, including the strength of evidence for individual effects within the health category. Further, U.S. EPA communicates what new insight the recent evidence provides, particularly where causal determinations have changed from the previous review. Although not addressed in this paper, U.S. EPA characterizes evidence relevant to understanding concentration-response relationships, exposure concentrations, durations, and patterns with which effects are observed, as well as populations and lifestages that may be at greater risk for effects [see Public Health Impact section of the Preamble to the ISAs (U.S. EPA 2015)].

2. DISCUSSION

ISAs inform policy decisions on the NAAQS. To ensure that decisions are based on a sound evaluation of the science, U.S. EPA uses a consistent and transparent framework to evaluate the causal nature of criteria pollutant-induced health effects in the ISA. The framework

shares features with other existing frameworks but is notable for emphasis on the integration of evidence across scientific disciplines and flexibility for application to diverse types of evidence and health effects. The standardized language in the framework was drawn primarily from EPA Cancer Guidelines (U.S. EPA 2005) and also from the National Academy of Sciences Institute of Medicine (IOM 2008), the International Agency for Research on Cancer (IARC 2006), and the Centers for Disease Control and Prevention (CDC 2004).

Rhomberg et al. (2013) reviewed about 50 weight-of-evidence frameworks, including the ISA framework, and highlights commonalities and distinguishing features. The surveyed frameworks use varying approaches, which is expected given they are designed for varying objectives that have unique challenges. Some frameworks, such as that by Adami et al. (2011), synthesize evidence from human epidemiologic and animal toxicological studies first by drawing conclusions for each and second by drawing overarching conclusions. Some frameworks evaluate specific components of risk assessment, such as mode of action (Sonich-Mullin et al. 2001; Boobis et al. 2006). In contrast, the ISA framework has broader purpose—to characterize relationships for many different pollutants and health effects from a varying mix of epidemiologic, controlled human exposure, and animal toxicological data. Rather than discipline-specific conclusions, the ISA framework emphasizes integration across disciplines and how evidence from one discipline can inform the interpretation of evidence from other disciplines. For example, confidence in evidence from one discipline may be low because there are few studies or uncertainty remains regarding potential confounding. But, these uncertainties may be sufficiently eliminated if there is coherence with findings from other disciplines for related outcomes (e.g., increased airway responsiveness and asthma symptoms). By filling data gaps, the evidence integrated across disciplines may provide the confidence to conclude that exposure to a pollutant could plausibly lead to a health effect.

Rhomberg et al. (2013) identified four phases of a weight-of-evidence process: “(1) define the causal question and develop criteria for study selection, (2) develop and apply criteria for review of individual studies, (3) integrate and evaluate evidence, and (4) draw conclusions based on inferences.” Goodman et al. (2013) compared these “best practices” to the ISA causal framework and recommended clarification on issues they stated are not explicitly described in the Preamble to the ISAs, namely, study selection criteria, study quality evaluation, Bradford Hill aspects application, and alternative hypotheses weighting. As described in the following paragraphs, U.S. EPA addresses the issues raised by Goodman et al. (2013) that may be specific to a particular pollutant or health effect in individual ISAs rather than in the Preamble.

Details on study identification and selection (e.g., dose cut-off, route of exposure) are included in introductory sections of an ISA or in the Integrated Review Plan, which is published at the start of a NAAQS review to describe the scope of the review and approach for developing the ISA. The approach for evaluating study quality, including specific considerations for a pollutant (e.g., potential confounding factors, exposure measurement error) has been described in an appendix to an ISA (U.S. EPA 2016). As illustrated in the examples below, the approach for evaluating study quality provides a basis for forming and

communicating judgments about strengths and uncertainties in the evidence base but does not serve as a strict checklist to define study quality. CASAC, which advises the U.S. EPA Administrator on the scientific basis for the NAAQS, expressed "some concern about applying strict evaluation criteria to various studies" and specific CASAC panelists advised U.S. EPA to "ensure these [guidelines] do not become checklists" (Diez and Frey 2015). U.S. EPA does not use formulaic approaches to causal determination and thus, does not quantitatively score study quality. Rather, the ISA describes strengths and limitations of studies to characterize the evidence and support judgments on the extent of error, bias, and strength of inference from results. Assessments of study quality are based on the judgments of ISA authors and are evaluated by CASAC as part of their review of a draft ISA.

Application of Bradford Hill aspects and weighting of alternative hypotheses are part of the discussion of the rationale for individual causal determinations in an ISA. Consideration of these issues can also be found in tables that accompany causal determinations and provide transparency to underlying judgments made on the weight and integration of evidence (example shown in Table I of Supplementary Material). Although strength of association was the first consideration in Bradford Hill's commentary (Hill 1965), consistency, coherence and biological plausibility figure most prominently in causal determinations for criteria pollutant-health effect relationships, which are generally characterized by small effect sizes. Small effect size can translate into a large impact for the population as a whole, and small effects can be reliably detected if a study is adequately powered and if important confounding factors are controlled. Further, large effects may be less plausible (Ioannidis 2016), for example, if important confounding factors are not considered. In evaluating consistency of the evidence, U.S. EPA focuses on the pattern of association across studies and not solely on statistical significance. Statistical significance is just one of the means to assess confidence in the observed relationship and the probability of chance as an explanation and is influenced by many factors such as the size of the study, exposure and outcome measurement error, and statistical model specifications (Greenland et al. 2016). In forming judgments on consistency, U.S. EPA also examines whether associations are observed across populations with varying risk factors and across locations with possibly varying pollutant mixtures.

2.1. Examples of Causal Determinations

The current framework was developed over many years as it has been applied to ISAs for each of the criteria pollutants, and it was recently published separately as a stand-alone document (U.S. EPA 2015). To demonstrate how the causal framework is implemented in the ISA, we highlight conclusions drawn in several recently finalized ISAs to show how judgments on study quality, Bradford Hill aspects, and alternative hypotheses inform causal determinations. The examples vary considerably in the nature of available evidence and the relative strength among disciplines and show how the causal framework was designed to be applicable to a broad range of evidence.

2.1.1. Causal Relationship—U.S. EPA concluded a "*causal relationship"* exists between a criteria pollutant and health effect in cases where experimental studies provided the primary evidence to rule out chance, confounding, and other biases with reasonable

confidence and in cases where epidemiologic studies provided key support. For the "*causal relationship"* determined between short-term O_3 exposure and respiratory effects in the 2013 ISA for Ozone, consistent evidence of O_3-induced lung function decrements in controlled human exposure studies provided the strongest support for an independent effect of O_3 exposure (U.S. EPA 2013a). Epidemiologic evidence was judged to be generally consistent and coherent, showing associations of O_3 with lung function decrements, pulmonary inflammation, as well as respiratory symptoms, emergency department visits, and hospital admissions. Although some results were not positive or statistically significant, the results showed a pattern of positive associations. For example, lung function decrements were linked to O_3 measured at the locations where subjects engaged in outdoor activity (e.g., summer camp), which has shown stronger correlations with measurements of personal O_3 exposure than O_3 measured at community central site monitors. Many studies observed that O_3 was associated with respiratory effects after statistical adjustment for fine particulate matter—a potentially confounding copollutant. However, the O_3-induced lung function decrements demonstrated in controlled human exposure studies and additional evidence from controlled human exposure and animal toxicological studies for O_3-induced increases in airway responsiveness, increases in allergic inflammation, and activation of neural reflexes gave U.S. EPA confidence that the epidemiologic results for O_3-associated respiratory effects could be attributable to O_3 exposure rather than a correlated copollutant. The design of the controlled human exposure and animal toxicological studies allowed assessment of the effects of O_3 exposure itself on the respiratory tract by exposing to O_3 alone, thus ruling out the confounding effects of other pollutants or other factors related to both O_3 and respiratory effects. Bias from the experimental procedures is reduced by using appropriate control exposures and methodology (e.g., randomization, double-blind exposures). Chance is ruled out, in part, by statistical significance of individual studies and also by the consistency of results among studies and across disciplines. The evidence integrated across scientific disciplines and health effects was judged sufficient to describe a biologically plausible pathway for short-term O_3 exposure to induce respiratory effects.

For the "*causal relationship*" determined for Pb exposure and cognitive function decrements in children in the 2013 ISA for Lead (U.S. EPA 2013b), epidemiologic evidence provided the strongest support. Several longitudinal studies following children from birth or infancy to adolescence observed associations between higher blood Pb level and lower intelligence quotient (IQ). Many of these studies were judged to be high quality. For one, the repeated measurement of blood Pb level and IQ supported the temporality of the relationship. Further, associations were observed with blood Pb levels measured in infancy or a very young age (e.g., 4 years). These associations were considered to better account for influences of higher blood Pb levels due to higher Pb exposures earlier in life. Findings of blood Pb-IQ associations with statistical adjustment for several potential confounders such as parental IQ and education, household income, race, parental caregiving quality, and smokers in the home supported an independent relationship with Pb exposure. A large consistent evidence base of animal toxicological studies provided biological plausibility for epidemiologic associations to rule out alternative explanations for the epidemiologic findings, including confounding by correlated risk factors such as low socioeconomic status or exposures to other metals. In particular, Pb-exposed rodents and monkeys showed impaired spatial memory and executive

function, effects also observed in epidemiologic studies. Animal toxicological studies demonstrating Pb effects on neuronal development, synaptic plasticity, and neurotransmitter function described a potential mode of action for Pb effects on cognitive function.

2.1.2. Likely to be a Causal Relationship—A determination of "*likely to be a causal relationship"* is appropriate when evidence shows a relationship between exposure to a pollutant and health effects but some uncertainties remain. Such was the case for short-term CO exposure and cardiovascular morbidity in the 2010 ISA for Carbon Monoxide (U.S. EPA 2010). Controlled human exposure studies using carboxyhemoglobin as a biomarker of CO exposure consistently demonstrated a CO-induced decrease in time to onset of angina and electrocardiogram changes at carboxyhemoglobin levels of 2–6%. Epidemiologic studies showed associations between ambient CO concentrations and cardiovascular hospitalizations that were generally robust to copollutant adjustment, but ambient CO concentrations are likely to result in smaller increases in carboxyhemoglobin levels than those produced endogenously and in controlled human exposure studies of CO. Thus, there was uncertainty as to the biological mechanism by which small carboxyhemoglobin changes could lead to effects observed in the epidemiologic studies. Unlike the examples provided for a "*causal relationship"*, cross-discipline integration could not address the uncertainty regarding biological plausibility for the cardiovascular effects associated with ambient CO exposures.

2.1.3. Suggestive of, but Not Sufficient to Infer, a Causal Relationship—A determination of "*suggestive of, but not sufficient to infer, a causal relationship"* does not require evidence demonstrating a cause-effect relationship. There is supporting evidence, but chance, confounding, and other biases cannot be ruled out. For short-term exposure to nitrogen dioxide (NO_2, the indicator for oxides of nitrogen) and cardiovascular effects, the 2016 ISA for Oxides of Nitrogen characterized the epidemiologic evidence base as large and consistent (U.S. EPA 2016). That is, several large time-series studies conducted over many years and in diverse locations associated short-term increases in ambient NO_2 concentrations with increases in emergency department visits and hospital admissions for myocardial infarction and ischemic heart disease as well as cardiovascular mortality. However, alternative explanations for the associations were not ruled out. Among the few studies that examined confounding by another traffic-related pollutant, the NO_2 association often became null with adjustment for fine particulate matter or CO. Experimental studies were inconsistent in showing NO_2–induced inflammation or oxidative stress, which could be early events leading to myocardial infarction. The evidence was determined "*suggestive of, but not sufficient to infer, a causal relationship"* because experimental studies had not demonstrated an independent effect of NO_2 exposure, and exposure to other traffic-related pollutants could not be ruled out as a plausible alternative explanation for the epidemiologic associations between NO_2 concentrations and cardiovascular effects.

As described in the 2013 ISA for Lead, few studies were available on Pb exposure and auditory function in adults, but a longitudinal study was judged to be high quality (U.S. EPA 2013b). In this study of healthy men, higher tibia bone Pb level was associated with a more rapid decrease in detection of sounds over a 23-year follow-up. The association was observed with adjustment for potential confounders, including age, race, education,

smoking, and occupational noise. Bone Pb level was measured near the end of follow-up, but temporality was not a major uncertainty as bone Pb is a good indicator of long-term Pb exposure. The other available epidemiologic study observed an association, but temporality was uncertain because auditory function and blood Pb levels were measured concurrently. Pb exposure of rodents and monkeys induced auditory function decrements, but the extent to which this evidence provided biological plausibility was unclear because Pb exposures were higher than those considered relevant to current ambient exposures. The evidence for Pb exposure and auditory function decrements in adults was considered "*suggestive of, but not sufficient to infer, a causal relationship"* given a strong epidemiologic study but uncertainty due to few studies overall and weak biological plausibility.

2.1.4. Inadequate to Infer a Causal Relationship—The determination of "*inadequate to infer a causal relationship"* is reached when the available studies are judged insufficient in quantity, quality, or consistency. For Pb exposure and visual function in children (U.S. EPA 2013b), the 2013 ISA for Lead notes that epidemiologic studies of visual impairments and animal toxicological studies with ambient-relevant Pb exposure concentrations were unavailable. In the case of long-term NO_2 exposure and postnatal development, the 2016 ISA for Oxides of Nitrogen notes a large epidemiologic evidence base (U.S. EPA 2016). However, results were inconsistent, including those which U.S. EPA judged to be based on good exposure estimates, residential or school exposures estimated with models that well captured the spatial variability of ambient NO_2 concentrations. Additionally, there was little biological plausibility as animal toxicological evidence was limited and inconsistent.

2.1.5. Not Likely to be a Causal Relationship—The determination of "*not likely to be a causal relationship"* is reached when several well-conducted studies show no effect across the full range of exposure levels and in potentially at-risk populations and lifestages. As characterized in the 2010 ISA for Carbon Monoxide, for long-term CO exposure and mortality, epidemiologic associations were predominantly negative or null (U.S. EPA 2010). One study (Lipfert et al. 2006) reported a positive association that disappeared when a variable representing traffic density was added to the model, suggesting another traffic pollutant was responsible for the association. Further, associations were absent among potential at-risk populations such as men with hypertension and post-menopausal women. These epidemiologic findings were coherent with results of studies showing neither evidence of, nor a potential mechanism for, effects of long-term CO exposure on respiratory and cardiovascular morbidity, the major causes of mortality.

3. CONCLUSIONS

Key concepts in U.S. EPA's approach to evaluate the causal nature of relationships between criteria air pollutants and health effects in the ISA are consistency, clarity, transparency, and flexibility. The framework provides standardized approaches and language to ensure consistency in evidence evaluation and causal determinations among health effects evaluated across ISAs. U.S. EPA emphasizes clarity and transparency in presentation of the evidence and characterization of weight-of-evidence conclusions used to draw each causal determination. As illustrated in the examples of causal determinations, flexibility is also

essential for application of the framework across a broad range of pollutants and health effects with varying types and availability of evidence. Differences between scientists in the interpretation of scientific evidence is expected; however, a consistent and well-described process for evidence evaluation based on scientifically-sound principles is essential for transparent discourse and review of the evidence in the ISA. The flexibility of U.S. EPA's causal framework for the ISA makes it a valuable tool for evaluations of evidence across media and health effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We gratefully acknowledge Tom Luben and Michael Stewart for their thoughtful reviews of this manuscript and past and current staff of the National Center for Environmental Assessment for their contributions to developing the Preamble to the Integrated Science Assessments. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abbreviations:

CASAC	Clean Air Scientific Advisory Committee
CO	carbon monoxide
IQ	intelligence quotient
ISA	Integrated Science Assessment
NAAQS	National Ambient Air Quality Standards
NO_2	nitrogen dioxide
O_3	ozone
Pb	lead
U.S. EPA	United States Environmental Protection Agency

REFERENCES

Adami HO, Berry SC, Breckenridge CB, Smith LL, Swenberg JA, Trichopoulos D, et al. 2011 Toxicology and epidemiology: Improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. Toxicological sciences: an official journal of the Society of Toxicology 122:223–234. [PubMed: 21561883]

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, et al. 2006 IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical reviews in toxicology 36:781–792. [PubMed: 17118728]

Clean Air Act. 1990a Clean Air Act, as amended by Pub. L. No. 101–549, section 108: Air quality criteria and control techniques, 42 USC 7408

Clean Air Act. 1990b Clean Air Act, as amended by Pub. L. No. 101–549, section 109: National primary and secondary ambient air quality standards, 42 USC 7409

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.

Diez R, Frey C. 2015 CASAC review of EPA's integrated science assessment for oxides of nitrogen – health criteria (second external review draft – January 2015) EPA-CASAC-15–001 Washington, DC: U.S. Environmental Protection Agency.

Goodman JE, Prueitt RL, Sax SN, Bailey LA, Rhomberg LR. 2013 Evaluation of the causal framework used for setting national ambient air quality standards. Critical reviews in toxicology 43:829–849. [PubMed: 24090029]

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. 2016 European Journal of Epidemiology 31:337–350. [PubMed: 27209009]

Hill AB. 1965 The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58:295–300. [PubMed: 14283879]

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

Ioannidis JP. 2016 Exposure-wide epidemiology: revisiting Bradford Hill. Statistics in medicine 35(11):1749–62. [PubMed: 26646432]

Institute of Medicine. 2008 Improving the presumptive disability decision-making process for veterans In Samet JM; Bodurow CC (Eds.). Washington, DC: National Academies Press.

Lipfert FW, Wyzga RE, Baty JD, Miller JP. 2006 Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects: Long-term mortality in a cohort of us veterans. Atmos Environ 40:154–169.

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, et al. 2013 A survey of frameworks for best practices in weight-of-evidence analyses. Critical reviews in toxicology 43:753–784. [PubMed: 24040995]

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, et al. 2001 IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regulatory toxicology and pharmacology: RTP 34:146–152. [PubMed: 11603957]

U.S. Environmental Protection Agency. 2005 Guidelines for carcinogen risk assessment EPA/630/P-03/001F Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum.

U.S. Environmental Protection Agency. 2008 Integrated Science Assessment for Oxides of Nitrogen – Health Criteria EPA/600/R-08/071 Washington, DC: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 2010 Integrated Science Assessment for Carbon Monoxide EPA/600/R-09/019F Washington, DC: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 2013a Integrated Science Assessment of Ozone and Related Photochemical Oxidants EPA/600/R-10/076F Washington, DC: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 2013b Integrated Science Assessment for Lead EPA/600/R-10/075F Washington, DC: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 2015 Preamble to the Integrated Science Assessments EPA/600/R-15/067 Washington, DC: U.S. Environmental Protection Agency [Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310244].

U.S. Environmental Protection Agency. 2016 Integrated Science Assessment for Oxides of Nitrogen – Health Criteria EPA/600/R-15/068 Washington, DC: U.S. Environmental Protection Agency.

Table 1

Aspects to aid in judging causality.

Aspect	Description
Consistency	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. Statistical significance is not the sole criterion by which the presence or absence of an effect is determined. 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, animal, or welfare studies) may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association. There may be coherence in demonstrating effects from evidence across various fields and/or across multiple study designs or related health endpoints within one scientific line of evidence.
Biological plausibility	An inference of causality is strengthened by results from experimental studies or other sources demonstrating biologically plausible mechanisms. A proposed mechanism, which is based on experimental evidence and which links exposure to an agent to a given effect, is an important source of support for causality.
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 or may not 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.
Temporality 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.

Source: Table I from U.S. EPA (2015) adapted from Hill (1965)

Table 2

Weight of evidence for causal determination in the U.S. EPA's Integrated Science Assessments.

Determination	Description of Weight of Evidence
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (e.g., doses or exposures generally within one to two orders of magnitude of recent concentrations). That is, the pollutant has been shown to result in health effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence. For example: (1) controlled human exposure studies that demonstrate consistent effects, or (2); observational studies that cannot be explained by plausible alternatives or that are supported by other lines of evidence (e.g., animal studies or mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups.
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, the pollutant has been shown to result in health effects in studies where results are not explained by chance, confounding, and other biases, but uncertainties remain in the evidence overall. For example: (1) 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 (2) animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited or no human data are available. Generally, the determination is based on multiple high-quality studies.
Suggestive of, but not sufficient to infer, a causal relationship	Evidence is suggestive of a causal relationship with relevant pollutant exposures but is limited, and chance, confounding, and other biases cannot be ruled out. For example: (1) when the body of evidence is relatively small, at least one high-quality epidemiologic study shows an association with a given health outcome and/or at least one high-quality toxicological study shows effects relevant to humans in animal species, or (2) when the body of evidence is relatively large, evidence from studies of varying quality is generally supportive but not entirely consistent, and there may be coherence across lines of evidence (e.g., animal studies or mode of action information) to support the determination.
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.
Not likely to be a causal relationship	Evidence indicates there is 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 and lifestages, are mutually consistent in not showing an effect at any level of exposure.

Source: Table II from U.S. EPA (2015).