

Comments on EPA's "Integrated Review Plan for the Primary National Ambient Air Quality Standards for Sulfur Dioxide, External Review Draft"

Prepared by Julie E. Goodman, Ph.D., DABT
on behalf of the
American Petroleum Institute
1220 L St NW
Washington, DC 20005

April 14, 2014



GRADIENT

www.gradientcorp.com
20 University Road
Cambridge, MA 02138
617-395-5000

Executive Summary

The United States Environmental Protection Agency (EPA) released the *Integrated Review Plan for the Primary National Ambient Air Quality Standards for Sulfur Dioxide (External Review Draft)* (draft IRP) in March 2014 (US EPA, 2014). The draft IRP reviews key findings from EPA's 2010 sulfur dioxide (SO₂) National Ambient Air Quality Standards evaluation and outlines its approach for the forthcoming Integrated Science Assessment (ISA) and Health Risk and Exposure Assessment (REA). Overall, the draft IRP presents a reasonable approach for conducting the ISA and REA. However, we identified several instances where the draft IRP could be more explicit and additional issues that EPA should consider or evaluate in the ISA and REA.

- In the draft IRP, EPA should be more explicit with regard to: 1) its literature search strategy; 2) how it will integrate studies it evaluated in the 2008 ISA (US EPA, 2008) with those it reviewed for the forthcoming ISA; 3) how it will determine study quality in a consistent manner across all studies; 4) how it will ensure accuracy of summarized information; and 5) how it will consider evidence suggesting a lack of effect of SO₂.
- In the ISA and REA, EPA should also consider: 1) whether observed effects in controlled human exposure studies are due to SO₂ or other factors; 2) evidence from all realms (*i.e.*, epidemiology, controlled exposure, animal toxicology) when interpreting results for individual studies; 3) uncertainties inherent in extrapolating potential health effects from high, experimental concentrations to low ambient exposure concentrations; 4) biological plausibility and coherence of proposed modes of action at ambient exposure levels; and 5) uncertainty in the risk estimates.
- Finally, EPA should revise its causal framework to more fully evaluate Bradford Hill's "aspects of association."

Introduction

The United States Environmental Protection Agency (EPA) released the *Integrated Review Plan for the Primary National Ambient Air Quality Standards for Sulfur Dioxide (External Review Draft)* (draft IRP) in March 2014 (US EPA, 2014). The draft IRP reviews key findings from EPA's 2010 sulfur dioxide (SO₂) National Ambient Air Quality Standards (NAAQS) evaluation (US EPA, 2010) and outlines its approach for the forthcoming Integrated Science Assessment (ISA) and Health Risk and Exposure Assessment (REA). EPA also released charge questions to the Clean Air Scientific Advisory Committee (CASAC) Sulfur Oxides (SO_x) Primary NAAQS Review Panel in March 2014 (CASAC Review Panel) (Sasser, 2014). These comments and, where relevant, responses to the accompanying charge questions address instances where the draft IRP should be more explicit and identify additional issues that EPA should consider in the ISA or REA.

Regarding Areas of Uncertainty in the Last Review (Section 3.1.3)

EPA should describe additional uncertainties from the 2010 review

In its charge questions to the CASAC Review Panel, EPA states, "Building on key considerations and issues addressed in the last review, Chapter 3 presents a set of policy-relevant questions that will serve as a focus in this review. To what extent does the Panel find that these questions appropriately characterize the key scientific and policy issues for consideration in the current review? Are there additional issues that should be considered?" (Sasser, 2014).

EPA describes five uncertainties from the 2010 review. Additional uncertainties that EPA should mention include 1) whether large percentage changes in measures of lung function, measured in forced expiratory volume in 1 second (FEV₁) and specific airway resistance (sRaw), in some individuals at low SO₂ concentrations (*i.e.*, 200-300 parts per billion, ppb) in controlled human exposure studies indicate an effect of SO₂ or, rather, are an artifact of intra-human variability in baseline values; and 2) whether populations exist that are more sensitive to SO₂ exposures than the mild-to-moderate exercising asthmatics in the controlled human exposure studies.

Regarding the first uncertainty, EPA should take variability in human lung function into account when evaluating the results of exposure to low SO₂ concentrations. For example, controlled human exposure studies measured lung function (*i.e.*, sRaw and/or FEV₁) in asthmatics before and after exercise in air containing 0 up to 1,000 ppb SO₂. At each SO₂ exposure level, the studies compared the percent change in lung function pre- to post-exercise with SO₂ exposure to the percent change in lung function pre- to post-exercise with no SO₂ exposure. While reported decrements were often statistically significant at SO₂ concentrations \geq 400 ppb, this was not the case at SO₂ concentrations of 200-300 ppb. EPA relies, however, on the 5-30% of individuals in these studies with apparent decrements in lung function at 200-300 ppb to assign an adverse effect of SO₂ at this level. Data from Bethel *et al.* (1985), which provided baseline (*i.e.*, pre-exercise and pre-exposure) sRaw values for 28 volunteers, demonstrate variability in human lung function: baseline sRaw measurements from two consecutive days varied by as much as 55%, and 11 of the 28 volunteers experienced a difference of 25% or greater. This variability in baseline values demonstrates that lung function can change appreciably due to factors unrelated to SO₂. In addition, variability in baseline values can lead to large percentage changes post-exposure, even if the exposure had no effect. Furthermore, in an evaluation of data from Linn *et al.* (1987), Linn (2010) showed that an equal number of individuals experienced lung function improvements and decrements after exercising at exposures of 200 ppb SO₂. Clinically relevant lung function decrements after exposure to 200 ppb SO₂ in a small number of individuals does not necessarily signify a clinically relevant adverse

effect from SO₂ exposure any more than lung function improvements in a small number of individuals signify a beneficial effect of SO₂. EPA could address the uncertainty of the effects of SO₂ at low concentrations in the controlled human exposure studies by using appropriate statistical analyses of group data rather than relying on selected individual-level data.

Regarding the second uncertainty, there is no evidence that a population exists that is more sensitive to SO₂ exposure than the volunteers in the available human exposure studies. The subjects in these studies represent asthmatics who exercise but do not take bronchodilators. Given that, for ethical reasons, controlled human exposure studies do not include individuals with the most severe asthma, it might be reasonably assumed, absent data to the contrary, that more severe asthma confers greater susceptibility to SO₂. As counterintuitive as it may seem, however, the available data do not support a lower threshold nor a larger response based on asthma severity. In their clinical trials, Linn *et al.* (1987) included volunteers classified into either "minimal/mild" or "moderate/severe" asthmatic groups. Within the range of SO₂ concentrations tested – 200, 400, and 600 ppb – the results do not support that asthma severity and SO₂ responsiveness are closely related. As further analyzed by Linn (2010), the Linn *et al.* (1987) data show "generally worse baseline status and generally worse effects of exercise in moderate/severe asthmatics, but generally similar incremental effects of SO₂ exposure in both groups." Thus, these data – from the only known available studies that address this issue – do not suggest that severe asthmatics need an extra margin of safety for protection against effects specific to SO₂. In addition, epidemiology studies provide no convincing evidence to suggest that children and elderly people are at an increased risk for respiratory effects at increased SO₂ levels.

Regarding the General Approach for the Current Review (Section 3.2)

The extent to which studies provide evidence that effects of SO₂ occur only at higher concentrations than previously thought should be included among the "key policy-relevant" issues

EPA rightfully plans to address the question, "To what extent has new information altered the scientific support for the occurrence of health effects as a result of short- and/or long-term exposure to sulfur oxides in the ambient air?" (US EPA, 2014, p. 3-14). A sub-bullet to this question, however, implies a bias toward finding effects of SO₂ at lower exposure concentrations: "Is there evidence of effects at exposure concentrations lower than have been previously observed or in areas that would likely meet the current SO₂ primary standard?" (US EPA, 2014). EPA should also consider scientific evidence that supports the occurrence of health effects from SO_x exposure only at *higher* concentrations than previously observed. For example, what if high-quality epidemiology studies have emerged that support a threshold for SO_x-related health effects that are higher than previously thought? Will EPA consider the merits of such studies for informing its overall weight-of-evidence (WoE) assessment? EPA should amend its General Approach to include consideration of all relevant data, not just those supporting a lower standard.

Regarding Literature Search and Selection of Relevant Studies (Section 4.3.2)

The draft IRP should provide more detail regarding its literature search strategy

In its charge questions to the CASAC Review Panel, EPA asks, "To what extent does Chapter 4 clearly and adequately describe the scope, approach, specific issues to be considered, and organization of the ISA? Please provide suggestions for any other issues that should be considered" (Sasser, 2014).

The draft IRP (US EPA, 2014) describes the literature search strategy but does not indicate which databases EPA will search, the complete search terms it will use, or specific study inclusion and exclusion criteria. The IRP should determine all of these factors *a priori*, with any changes discussed in the ISA. EPA should consider closely following the Cochrane Review process, which serves as a basis for several WoE frameworks, such as the one being developed by the Office of Health Assessment and Translation within the National Toxicology Program (Higgins and Green, 2011; NTP, 2012; Thayer, 2012; Boyles, 2012). A Cochrane Review is a systematic review of original studies of health care data (predominantly randomized and clinical controlled trials, but also observational studies) that uses a predefined, rigorous, and explicit methodology. The key goals of a Cochrane Review are to 1) collate all evidence that fits pre-specified eligibility criteria to address a specific research question; 2) minimize bias by using explicit, systematic methods; and 3) prepare, maintain, and promote systematic reviews to inform health care decisions. Specifically, investigators synthesize the primary results using strategies that limit bias and random error. These strategies include a comprehensive search of all potentially relevant studies and use of explicit, reproducible criteria in the selection of studies for review. Investigators appraise primary research designs and study characteristics, synthesize data, and interpret results. The *Cochrane Handbook for Systematic Reviews of Interventions*¹ describes the process of preparing and maintaining Cochrane Reviews in detail.

The draft IRP should clarify how it will integrate studies evaluated in the 2008 SO_x ISA with those reviewed for the forthcoming ISA

The draft IRP (US EPA, 2014, p. 4-5) states:

The ISA will generally emphasize studies published since the 2008 SO_x ISA; however, evidence from previous studies will be included to integrate with results from recent studies and, in some cases, characterize the key policy-relevant information in a particular subject area.

EPA should carefully consider and clearly articulate how it will incorporate evidence from older studies in the overall causality assessment. By incorporating older studies after the evaluation of recent studies, older studies may receive less weight in the analysis because of publication date. EPA should reanalyze the key older studies so that it considers them in the same manner as the newer studies.

Regarding Evaluation of Individual Study Quality (Section 4.3.3)

The draft IRP should provide more detail regarding how it will determine study quality, and how it will do so in a consistent manner, for all studies

The draft IRP (US EPA, 2014) notes that EPA will evaluate study quality by assessing such things as the representativeness of the exposure assessment, adequacy of the study population, appropriateness of statistical analyses, control of potential confounders, and validity and reliability of health endpoints. Throughout the plan, the draft IRP lists several important questions it will address in the ISA. All of these are crucial to consider when evaluating study quality and relevance.

The draft IRP should specify criteria for assessing these metrics and ensure they are applied in a consistent manner across studies. For example, before evaluating any study results, the ISA should discuss all of the ways in which exposure can be measured, the strengths and limitations of each method,

¹ <http://www.cochrane-handbook.org/>.

the possibility for exposure measurement error, and which methods carry the most weight. There should also be a discussion of statistical methods used among all studies evaluated and which specific methods are more robust and why (*e.g.*, whether multiple comparisons have been addressed and assumptions in Cox proportional hazard model are appropriate). The ISA should address specific confounders [*e.g.*, co-pollutants, socioeconomic status (SES), age, weather] in terms of how they are handled in different studies and their likely impact on results. Other factors the ISA should consider in detail include measurement bias, measurement precision, replicability of observations, data reliability, outliers, selective outcome reporting, and fraudulent studies.

It is crucial that EPA evaluate these quality measurements in the same way across studies. For example, if EPA considers a particular statistical model a limitation for one study, it must consider it a limitation in all studies that use the same model (unless there is a reason to conclude otherwise; if this is the case, EPA should state the reason clearly). Importantly, an evaluation of study quality should be independent of the results and funding source of that study – it should be based purely on the methods, in a consistent manner across studies. Further, studies with more robust methods should receive more weight in causal determinations. Currently, the draft IRP provides no explicit rationale for why certain studies are considered key evidence while others of similar quality are not.

It is expected that, during the review, EPA will find study quality criteria not determined *a priori*. If so, it should apply these criteria to earlier studies to ensure that the assessment is consistent across studies. By more precisely defining and applying these criteria, the ISA will be more transparent and balanced.

The ability to identify whether lung function effects are due to SO₂ or to other factors should be considered when evaluating study quality of controlled human exposure studies

As described above, day-to-day variability in baseline measures of FEV₁ and sRaw in controlled human exposure studies can lead to large percentage changes post-exposure that are not related to SO₂. This can lead to the erroneous attribution of lung function decrements to SO₂ in some individuals, especially at low concentrations (*e.g.*, 200 ppb SO₂). Given the importance of this potential source of uncertainty, EPA should consider the extent to which studies can address it (*e.g.*, by collecting multiple baseline measures per individual) when evaluating individual studies. At the very least, the ISA should consider that lung function decrements in individuals could be artifacts of intra-human baseline variability – rather than an effect caused by SO₂ – and rely instead on conducting appropriate statistical analyses of group data.

The ISA should evaluate *in vitro* studies that both support and call into question results demonstrated *in vivo*

The draft IRP states, "*In vitro* studies may be included if they provide mechanistic insight or support results demonstrated *in vivo*." Notwithstanding their limitations, the ISA should include *in vitro* studies not only if they support *in vivo* results, but also if they call *in vivo* results into question. In addition, the ISA should evaluate concordance of results from *in vitro* studies with results from epidemiology studies, as findings from epidemiology studies can often suffer from issues related to chance, bias, and/or confounding. In sum, the ISA should consider results from all relevant *in vitro* studies along with other lines of evidence, and it should determine whether (if they are of sufficient quality) the studies support or refute findings from epidemiology, controlled human exposure, or animal studies.

Regarding Integration of Evidence and Determination of Causality (Section 4.3.4)

EPA's NAAQS causal framework should be revised

The NAAQS causal framework incorporates language from sources across the federal government and scientific community, particularly the Institute of Medicine (IOM) report *Improving the Presumptive Disability Decision-making Process for Veterans* (IOM, 2008). Whereas the IOM recommended four categories for the level of evidence for causation (Table 1), the NAAQS has five categories for causal relationships (Table 2). Based on these categories, EPA determines which health effects will be evaluated in quantitative risk assessments (US EPA, 2013).

EPA's causal framework is largely based on modified Bradford Hill aspects. Both the original and modified Bradford Hill aspects (*i.e.*, strength of association, consistency and coherence, biological plausibility, biological gradient or exposure-response, specificity, temporality of effect, and adversity) are useful tools for evaluating causation. This is because it may be difficult to ascribe observations to causation if these aspects are not met, whereas it may be difficult to ascribe observations to anything other than causation if they are met. As discussed in Goodman *et al.* (2013), EPA's application of the causal framework is not congruent with the judgments based on the original or modified Bradford Hill aspects. For example, the framework claims to rely heavily on the aspect of consistency across studies in its categorization scheme, but, in practice, it does not always fully evaluate consistency or incorporate aspects such as coherence, biological plausibility, biological gradient, and strength of association.

Based on the current NAAQS causal framework, the determination of a causal relationship cannot be made reliably without fully exploring chance, bias, and confounding, but the IRP should make evident that the ISA will do this in a consistent manner. Furthermore, the IRP should indicate that the ISA should fully explore whether and to what degree the data support hypotheses *other* than the criteria pollutant causes a particular health effect (*e.g.*, a plausible confounder, rather than the criteria pollutant, causes a particular health effect). It is only in this manner that one can truly explore alternative hypotheses.

For making determinations regarding causality, it is important to evaluate all available data, including positive, null, and negative evidence, in a WoE evaluation. Any WoE evaluation, by definition, involves a consideration of all lines of evidence in a consistent manner. It is not about resolving all uncertainty but, rather, determining whether the evidence as a whole supports causation more than it supports a lack of effect. If co-pollutants cannot be addressed or studies are inconsistent, the WoE may indicate a lack of causality or inadequate evidence to assess causation. If positive effects in high-dose animal studies cannot be related to humans, this does not constitute suggestive evidence; instead, these effects are essentially uninformative regarding causation in humans. Not every study evaluating a criteria pollutant is informative for evaluating human health risk, and the ISA should not place undue weight on studies that are not.

It is notable that the NAAQS causal framework requires only one high-quality study for evidence of causal relationship to be deemed *suggestive*. Using this definition, high-quality studies that are inconsistent with evidence of an association may exist but – as long as one high-quality study demonstrates an effect – there would still be enough evidence to constitute a suggestive relationship. Instead, all studies should be reviewed using the same criteria and one should conclude a suggestive causal association only if the WoE indicates that a causal association is more likely than not based on all the data combined. In situations where there are multiple, but inconsistent, high-quality studies, the appropriate conclusion is that the evidence is inadequate to determine causality. Because of this issue, we

recommend eliminating the suggestive category; EPA should consider endpoints in this category either above or below equipoise, as suggested in the IOM (2008) framework.

Finally, evaluating the evidence as a whole means that one should evaluate not only how much evidence can be adduced to support (or to counter) the hypothesized causal effect, but how separate lines of evidence support (or contradict) one another. That is, it is critical to determine the most likely explanation for discrepancies across studies by evaluating all of the evidence and not selectively considering data that supports or counters a given hypothesis.

Many of the issues noted above could be resolved if the EPA revises the NAAQS framework to make categories for causal determination more similar to the IOM framework on which it was based originally. The ISA should evaluate all evidence in a consistent manner using well-specified criteria and determine whether, as a whole, it constitutes evidence for causation or is more likely indicative of an alternative hypothesis. EPA should proceed with a risk assessment on a particular health effect only if the evidence is clearly supportive of causation (*i.e.*, equipoise and above in the IOM framework).

The ISA should consider each realm of evidence when interpreting study results

The NAAQS causal framework looks separately at epidemiology, controlled exposure, and animal toxicology evidence, first coming to a synthesized judgment for each, and then integrating these separate judgments into an overall qualitative statement about causality (US EPA, 2013). As discussed by Goodman *et al.* (2013), the data evaluation should be integrated across all lines of evidence *before* coming to judgments based on each realm independently. In this way, interpretation of each line of evidence informs the interpretation of the others. For example, if one can interpret an epidemiology analysis in two different ways, and animal studies can shed light on whether one is more plausible than the other, this should be considered when making judgments about the epidemiology study.

Regarding Quality Management (Section 4.3.5)

The draft IRP should be explicit regarding quality assurance measures that will be used to ensure accuracy of summarized information

The draft IRP (US EPA, 2014) states, "Where information is integrated, re-analyzed, modeled, or reduced from multiple sources to create new figures, tables, or summation, the data generated are considered to be new and are documented and subjected to rigorous quality assurance and quality control measures to ensure their accuracy, validity, and reproducibility." The draft IRP should provide additional detail regarding the quality assurance and quality control measures; for example, will information be extracted and summarized by two different individuals and compared for accuracy, or will information extracted and summarized by one individual be checked by a second individual? In addition to accuracy, the ISA should ensure that information is not extracted or summarized in a selective manner.

Regarding Specific Issues to be Addressed in the ISA (Section 4.4)

The draft IRP should be more explicit with respect to evidence required for determining a lack of effect

The draft IRP (US EPA, 2014) indicates that two issues to be addressed in the ISA are "whether new evidence reinforces or calls into question the evidence presented and evaluated in the last NAAQS review with respect to factors such as the concentrations of SO_x exposure associated with health effects and plausibility of health effects caused by exposure to SO_x exposure" and "whether uncertainties from the last review have been reduced and/or whether new uncertainties have emerged."

EPA needs to describe how it will determine when evidence calls a causal association into question. For example, it is often the case that evidence indicates a lack of causation to be as likely, or even more likely, than causation (*e.g.*, if confounders cannot be totally accounted for or if exposure misclassification causes false positive results). There is a tendency to conclude that because of the *possibility* for causation, the data supports causality. Instead, in this case, one should conclude the evidence is non-informative.

If similar studies come to similar conclusions but all indicate that lack of causation is as likely as causation, they should not contribute to a causal conclusion. As stated by Bradford Hill (1965), "the same results from precisely the same form of inquiry will not invariably strengthen the original evidence. I would myself put a good deal of weight upon similar results reached in quite different ways."

The draft IRP should state that the ISA will only conclude that an association is causal or likely causal if such an association is *more* likely than not.

EPA's "at-risk" factors should more appropriately be referred to as effect modifiers

The draft IRP states that the ISA "will examine exposure and health outcome data to draw conclusions about specific populations or lifestages that are potentially at increased risk of SO_x-induced health effects" (US EPA, 2014, p. 4-16). It describes that potential "at-risk" populations or lifestages can be characterized by a variety of factors, including both intrinsic (biological factors such as age and genetic variants), extrinsic (nonbiological factors such as diet and lower SES), and/or other factors affecting dose or exposure (*e.g.*, age, outdoor activity, or work). An "at-risk" factor is more accurately described as an effect modifier, which is a technical term defined in epidemiology as a variable that differentially modifies the observed effect of a risk factor (in this case, SO_x) on disease status. To the extent that "at-risk" factors are evaluated in the ISA based on epidemiology studies, it may be more appropriate for EPA to refer to them as effect modifiers, as we do herein.

EPA's frameworks for both causal determination and classification of evidence for potential effect modifiers are based on consideration of modified Bradford Hill aspects, but the former has five categories and the latter four and the criteria for classification are not the same. It appears that, although defined differently, the four categories in the framework for effect modifiers are roughly equivalent to *causal relationship*, *suggestive of a causal relationship*, *inadequate to infer a causal relationship*, and *not likely to be a causal relationship* in the NAAQS causal framework; there does not appear to be an equivalent to the *likely to be a causal relationship*. Although one is an assessment of direct causation and the other an assessment of factors that can contribute to (or prevent) causation, in both cases, the goal is to critically, systematically, and transparently review the weight of scientific evidence. Ideally, the same rules should be applied for both types of analysis; if not, there needs to be justification for using different rules to

conduct the same type of analysis. The ISA should adopt the IOM-recommended categories for the level of evidence for causation, which consider whether the WoE is above or below equipoise (IOM, 2008).

Issues with the WoE for causal determination apply to the effect modifier classifications as well (see Goodman *et al.*, 2013). For example, the ISA defines evidence for an effect modifier to be suggestive if it "is limited due to some inconsistency within a discipline or, where applicable, a lack of coherence across disciplines." If an inconsistency or a lack of coherence is large enough, it should lead to a conclusion that the WoE is below equipoise. In these circumstances, the evidence is inadequate to make a determination as to whether a factor is an effect modifier for a health endpoint.

Regarding Consideration of Quantitative Assessments for This Review (Section 5.2)

Uncertainty should be addressed in the risk estimates

In its charge questions to the CASAC Review Panel, EPA asks, "To what extent does Chapter 5 clearly and adequately describe the scope and specific issues, including the identification of the most important uncertainties, to be considered in developing the REA Planning Document for this review? To what extent is there additional information that should be considered or additional issues that should be addressed in considering the potential for risk and/or exposure analyses in the current review?" (Sasser, 2014).

To reduce uncertainty overall, quantitative risk estimates should focus on endpoints for which there is strong evidence of causal association, with robust data for both air quality evaluations and concentration-response functions. To the extent possible, evaluations in the REA should incorporate quantitative estimates of uncertainty into the confidence bounds around risk estimates. If it is not possible to quantify certain aspects of uncertainty, the REA should indicate whether the uncertainty is likely to over- or underestimate risks and provide a qualitative indication of the magnitude of the uncertainty (*e.g.*, high, medium, or low). The draft IRP should indicate that, if an aspect of uncertainty could produce both outcomes, the REA should provide examples of how and when the uncertainty would under- or overestimate risks.

Conclusions

Although the draft IRP presents a reasonable approach for conducting the ISA and REA, there are some instances where the draft IRP could be more explicit and other instances where it should consider or evaluate additional issues for conducting the ISA or REA. In addition, the EPA's NAAQS causal framework should be revised to more fully evaluate Bradford Hill's "aspects of association."

Table 1 IOM Recommended Categories for the Level of Evidence for Causation

Causal Determination	Evidence
Sufficient	The evidence is sufficient to conclude that a causal relationship exists. For example: a) replicated and consistent evidence of an association from several high-quality epidemiologic studies that cannot be explained by plausible noncausal alternatives (<i>e.g.</i> , chance, bias, or confounding); or b) evidence of causation from animal studies and mechanistic knowledge; or c) compelling evidence from animal studies and strong mechanistic evidence from studies in exposed humans, consistent with (<i>i.e.</i> , not contradicted by) the epidemiologic evidence.
Equipoise and above	The evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists. For example: a) evidence of an association from the preponderance of several high-quality epidemiologic studies that cannot be explained by plausible noncausal alternatives (<i>e.g.</i> , chance, bias, or confounding) as well as animal evidence and biological knowledge consistent with a causal relationship; or b) strong evidence from animal studies or mechanistic evidence that is not contradicted by human or other evidence.
Below equipoise	The evidence is not sufficient to conclude that a causal relationship is at least as likely as not, or is not sufficient to make a scientifically informed judgment. For example: a) consistent human evidence of an association that is limited by the inability to rule out chance, bias, or confounding with confidence, and weak animal or mechanistic evidence; or b) animal evidence suggestive of a causal relationship, but weak or inconsistent human and mechanistic evidence; or c) mechanistic evidence suggestive of a causal relationship, but weak or inconsistent animal and human evidence; or d) the evidence base is very thin.
Against	The evidence suggests the lack of a causal relationship. For example: a) consistent human evidence of no causal association from multiple studies covering the full range of exposures encountered by humans; or b) animal or mechanistic evidence supportive of a lack of a causal relationship.

Source: IOM (2008).

Table 2 EPA's Weight of Evidence for Causal Determination

Causal Determination	Health Effects
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (<i>i.e.</i> , 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 (<i>e.g.</i> , animal studies or mode of action information). Evidence includes multiple high-quality studies.
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.
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.
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 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.

Source: US EPA (2013, Table II).

References

- Bethel, RA; Sheppard, D; Geffroy, B; Tam, E; Nadel, JA; Boushey, HA. 1985. "Effect of 0.25 ppm sulfur dioxide on airway resistance in freely breathing, heavily exercising, asthmatic subjects." *Am. Rev. Respir. Dis.* 131(4):659-661.
- Boyles, AL. 2012. "NTP Update: Systematic Review Incorporating Data Mining with Graphical Tools for Researchers." Office of Health Assessment and Translation. 30p., March 6.
- Goodman, JE; Prueitt, RL; Sax, SN; Bailey, LA; Rhomberg, LR. 2013. "Evaluation of the causal framework used for setting National Ambient Air Quality Standards." *Crit. Rev. Toxicol.* 43(10):829-849.
- Higgins, JPT; Green, S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)*. The Cochrane Collaboration. Accessed at <http://www.cochrane-handbook.org/>, March.
- Hill, AB. 1965. "The environment and disease: Association or causation?" *Proc. R. Soc. Med.* 58:295-300.
- Institute of Medicine (IOM). 2008. "Improving the Presumptive Disability Decision-Making Process for Veterans." Committee on Evaluation of the Presumptive Disability Decision-Making Process for Veterans, Board on Military and Veterans Health, National Academies Press. Accessed at http://books.nap.edu/openbook.php?record_id=11908&_r=781p.
- Linn, WS. 2010. "A Critical Review of Selected Documents Concerning Proposed Revision of the National Ambient Air Quality Standard for Sulfur Dioxide." Report to American Petroleum Institute. 12p., January.
- Linn, WS; Avol, EL; Peng, RC; Shamoo, DA; Hackney, JD. 1987. "Replicated dose-response study of sulfur dioxide effects in normal, atopic, and asthmatic volunteers." *Am. Rev. Respir. Dis.* 136(5):1127-1134.
- National Toxicology Program (NTP). 2012. "Summary Minutes of NTP Board of Scientific Counselors Meeting, held June 21-22, 2012, in Research Triangle Park, North Carolina." 36p., June.
- Sasser, E. [US EPA, Office of Air Quality Planning and Standards]. 2014. Memorandum to D. Wong [US EPA, Clean Air Scientific Advisory Committee (CASAC)] re: CASAC Review of the "Draft Integrated Review Plan for the Primary National Ambient Air Quality Standard for Sulfur Dioxide." Accessed at [http://yosemite.epa.gov/sab/sabproduct.nsf/0/75696575A41DDF1985257B970066DB2B/\\$File/Draft%20IRP-CASAC%20Transmittal%20SO2%20\(3-18-2014\)%20\(2\).pdf](http://yosemite.epa.gov/sab/sabproduct.nsf/0/75696575A41DDF1985257B970066DB2B/$File/Draft%20IRP-CASAC%20Transmittal%20SO2%20(3-18-2014)%20(2).pdf), 4p., March 18.
- Thayer, K. 2012. "Systematic Review and New Tools for Information Management and Data Display." Office of Health Assessment and Translation. 1p.
- US EPA. 2008. "Integrated Science Assessment for Sulfur Oxides - Health Criteria." Office of Research and Development, EPA/600/R-08/047F. September.
- US EPA. 2010. "Primary national ambient air quality standard for sulfur dioxide (Final rule)." *Fed. Reg.* 75(119):35520-35603. June 22.

US EPA. 2013. "Integrated Science Assessment for Oxides of Nitrogen - Health Criteria (First External Review Draft)." National Center for Environmental Assessment, EPA/600/R-13/202. 878p., November.

US EPA. 2014. "Integrated Review Plan for the Primary National Ambient Air Quality Standards for Sulfur Dioxide (External Review Draft)." National Center for Environmental Assessment, Office of Air Quality Planning and Standards, EPA-452/P-14-005. March.