 DOI: 10.1111/risa.12405

Rethinking Meta-Analysis: Applications for Air Pollution Data and Beyond

Julie E. Goodman,[1,*] Catherine Petito Boyce,[2] Sonja N. Sax,[3] Leslie A. Beyer,[3] and Robyn L. Prueitt[2]

Meta-analyses offer a rigorous and transparent systematic framework for synthesizing data that can be used for a wide range of research areas, study designs, and data types. Both the outcome of meta-analyses and the meta-analysis process itself can yield useful insights for answering scientific questions and making policy decisions. Development of the National Ambient Air Quality Standards illustrates many potential applications of meta-analysis. These applications demonstrate the strengths and limitations of meta-analysis, issues that arise in various data realms, how meta-analysis design choices can influence interpretation of results, and how meta-analysis can be used to address bias and heterogeneity. Reviewing available data from a meta-analysis perspective can provide a useful framework and impetus for identifying and refining strategies for future research. Moreover, increased pervasiveness of a meta-analysis mindset—focusing on how the pieces of the research puzzle fit together—would benefit scientific research and data syntheses regardless of whether or not a quantitative meta-analysis is undertaken. While an individual meta-analysis can only synthesize studies addressing the same research question, the results of separate meta-analyses can be combined to address a question encompassing multiple data types. This observation applies to any scientific or policy area where information from a variety of disciplines must be considered to address a broader research question.

KEY WORDS: Air pollutants; bias; data synthesis; heterogeneity; meta-analysis

1. BACKGROUND

A meta-analysis is a type of systematic review that can be a powerful tool for assembling, critically appraising, and synthesizing data from multiple individual studies. Meta-analysis offers quantitative methods for combining multiple data sets addressing a specific research question to yield an overall "consensus" of the data.[(1)] Commonly used meta-analysis techniques include descriptive tables, graphical analyses, and statistical approaches. For example, a descriptive table can present the point estimate and standard error or confidence interval of the outcome measure from each study, making it easier to compare and detect patterns in the results. Similarly, results can be displayed and compared graphically, e.g., by preparing a forest plot. In meta-analyses, study results can also be weighted using various standards, e.g., relative study size, variation among observations, or factors influencing study quality.

Classical statistical methods that can be used to quantitatively synthesize study results include weighted averages and meta-regression, a tool that uses regression techniques to simultaneously examine how study results are affected by multiple potential modifying factors (including both continuous and categorical factors). Heterogeneity reflected in the study results under consideration can be formally

[1]Gradient, 20 University Rd., Ste. 5, Cambridge, MA 02138, USA.
[2]Gradient, Seattle, WA.
[3]Gradient, Cambridge, MA.
*Address correspondence to Julie E. Goodman, Gradient, 20 University Rd., Ste. 5, Cambridge, MA 02138, USA; tel: 617-395-5525; fax: 617-395-5001; jgoodman@gradientcorp.com.

 0272-4332/15/0100-0001$22.00/1 © 2015 Society for Risk Analysis

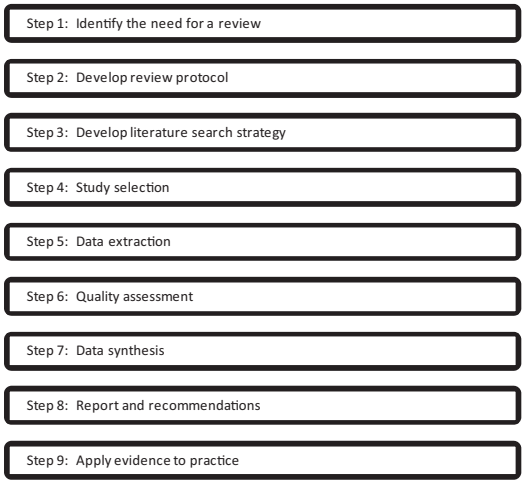

Fig. 1. Steps of a meta-analysis. Adapted from CRD.[8]

assessed using statistical tests such as Cochrane's Q statistic and Higgins's I^2 test.[2,3] When the study results are deemed to be adequately homogeneous, meta-analysis and meta-regression can be conducted using a fixed-effect model (which treats the effect size parameters as fixed values reflecting one true population effect). When the results are insufficiently homogeneous, a random-effects model can be used (which treats the effect size parameters from the studies as random values from some underlying population of values such as a normal distribution). Additional details regarding methods and issues in conducting meta-analyses and meta-regressions can be found in resources such as Blair *et al.*,[4] Nordmann *et al.*,[5] Rothman *et al.*,[6] and Stroup *et al.*[7]

A well-conducted meta-analysis prepared following the steps illustrated in Fig. 1 incorporates key features that help minimize bias, random errors, and subjectivity when evaluating data.[8] These features include requirements for: (1) a thorough literature search; (2) clear and transparent study eligibility criteria; (3) a standardized approach for critically appraising studies; (4) appropriate statistical calculations to assess comparisons and trends among study findings; and (5) evaluations of potential sources of bias and heterogeneity (of both study methods and results). While some of these features can be incorporated into more qualitative, narrative systematic study reviews (reviewed by Rhomberg *et al.*[9]), the more rigorous, quantitative perspective inherent in the meta-analysis approach can foster a more in-depth evaluation of study results and the factors that influence findings. Although a single meta-analysis can only evaluate studies of similar design addressing a specific research question, the methodology is adaptable to a wide range of research areas, study designs, and data types. Thus, it has the potential to play a valuable role in settings that must synthesize data from multiple disciplines.

Data from individual studies can be combined to form one data set and analyzed in what is called a pooled analysis, but the original data from individual studies often are not readily available to other researchers. A meta-analysis provides a way to combine results from individual studies when primary data are not available. Relative to the individual studies comprising the meta-analysis, the greater statistical power of the combined data can yield a more precise estimate of the outcome being studied, reduce the possibility of false negative results, provide evidence regarding potential study biases, and generate insights regarding sources of observed heterogeneity or other patterns in study results.[4] In addition, combining results of individual studies can make them more generally applicable (e.g., across various populations).[5] Overall, a soundly conducted meta-analysis can help researchers understand and reconcile apparent contradictions in study data (e.g., where available studies report positive and negative outcomes for the same endpoint).

Development of the National Ambient Air Quality Standards (NAAQS[10]) illustrates many potential applications of meta-analysis. Applying data from a variety of disciplines (including epidemiology, toxicology, atmospheric science, and exposure science), US EPA has developed NAAQS for six "criteria" air pollutants: carbon monoxide (CO), lead, nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM), and sulfur dioxide (SO_2). As part of its most recent NAAQS review process, US EPA is using a modified Bradford Hill framework to characterize the weight evidence (WoE) for causal determinations for each air pollutant and various human health and ecological outcomes.[4] For each substance, US EPA documents these evaluations in an Integrated Science Assessment (ISA) report.

[4]US EPA's framework adopts the nine "aspects" of study findings to be considered in judging causality identified in Hill;[20] e.g., strength, consistency, and plausibility of observed effects. Because the original formulation focused on epidemiology study results, US EPA modified the discussion of the aspects to make the framework more broadly applicable to other types of data, including results from controlled human exposure, ecological, animal toxicology, and *in vitro* studies (e.g., as summarized in Table I of US EPA[12]).

Final ISAs have been completed for four of the criteria pollutants: CO,[11] lead,[12] O_3,[13] and PM.[14] The current review processes for NO_2[15,16] and SO_2[17] are in the early stages.

US EPA and others have used meta-analyses to assess a limited amount of the data supporting the NAAQS determinations—primarily data from epidemiology and controlled exposure studies (see, e.g., US EPA;[18] Goodman *et al.*[19]). However, other types of supporting data (e.g., toxicology and mechanistic data) have not been evaluated routinely using meta-analysis, and additional opportunities exist to use this methodology. Drawing upon relevant examples from air pollutant research, this article discusses how meta-analysis has been used to integrate results from individual studies within specific research areas (e.g., studies addressing a specific health endpoint), with a focus on identifying innovative applications of the methodology. In particular, we examine the strengths and limitations of existing meta-analyses and identify opportunities to refine them or expand use of this methodology to other data types. Although our focus is on air pollutant evaluations, we also address (1) when and with what kinds of data meta-analyses can be useful across a variety of disciplines, (2) the implications of certain design choices on meta-analysis results, and (3) how meta-analysis can help when designing and implementing new research efforts. We also discuss how the results of separate meta-analyses can be combined to address a larger question (e.g., a causation determination) where findings from a variety of data types and research areas must be integrated.

2. CONTROLLED EXPOSURE STUDIES

In controlled exposure studies, people with regulated activity levels experience known concentrations of substances, such as air pollutants, in exposure chambers with carefully controlled conditions.[18] This exposure method minimizes possible confounding by other factors, and sensitive experimental techniques can be used to measure endpoints of interest that are generally not evaluated in observational epidemiology studies. The types of effects commonly studied in controlled exposure studies include reversible, acute effects from short-duration exposures. Typically, these effects are easily measured and can be described as categorical or continuous variables.[21] Controlled exposure studies often provide important information on health effects, quantitative exposure-response relationships, and the biological plausibility of associations identified in observational studies, as well as insights regarding sensitive subpopulations.

Although controlled exposure studies, by definition, allow for substantial control over experimental study conditions, such studies have a number of features that affect study interpretation.[18] First, subjects must be healthy enough to participate in the study, and the health effects evaluated must be transient, reversible, and not severe. Therefore, the results may underestimate health impacts for certain sensitive subpopulations and will not reflect effects that are persistent or occur following chronic exposures. Second, these studies often use concentrations that are higher than those normally present in ambient air, so any observed effects may not occur at the lower concentrations people typically experience. Third, these studies generally are conducted on a relatively small number of subjects, reducing the power of each study to detect statistically significant differences in the health outcomes of interest. Despite these limitations, controlled exposure studies are generally good candidates for meta-analysis because they often have homogeneous study designs and address the same research question. In fact, the small number of subjects per study makes meta-analysis a very important tool to evaluate these studies as a whole because meta-analysis can provide a systematic framework for combining the results from many small studies, increasing the statistical power to observe associations.

The current NAAQS determinations for O_3, SO_2, and NO_2 are based largely on controlled exposure study results, in addition to results from observational epidemiology studies. To date, NO_2 is the only criteria pollutant for which US EPA has conducted a meta-analysis of controlled exposure studies. In its 2008 *Integrated Science Assessment of Oxides of Nitrogen*, US EPA evaluated the effects on airway responsiveness to nonspecific challenge agents (e.g., carbechol, cold-dry air, histamine, methacholine) following NO_2 exposure in people with mild asthma.[18] US EPA classified individuals as having either an increase or decrease in airway response based on one of three measures (i.e., specific airway conductance (sGaw), specific airway resistance (sRaw), and forced expiratory volume in 1 s (FEV_1)). US EPA then conducted a meta-analysis to determine whether the percentage of people with lung function decrements was greater if they were exposed to NO_2 or filtered air.[18]

Altogether, the analysis included data from 17 studies of 355 asthmatics with one-hour exposures

ranging from 0.1 to 0.6 ppm NO_2. US EPA evaluated the combined data and data stratified by exposure level (0.1, 0.1–0.2, 0.2–0.3, and >0.3 ppm) and activity level (rest vs. exercise). US EPA reported that NO_2 was not associated with airway hyperresponsiveness (AHR) in people exposed while exercising, but it was associated with AHR in people exposed at rest at all exposure levels (although exposure-response was not evaluated specifically). US EPA's overall conclusion was that 60-min exposures to ≥0.1 ppm NO_2 were associated with small but significant increases in nonspecific AHR in people with mild asthma. In the Final Rule for NO_2, US EPA[22] stated that it was appropriate to consider NO_2-induced AHR in characterizing NO_2-associated health risks. It based the one-hour NAAQS of 0.1 ppm largely on this analysis, its evaluation of the observational epidemiology literature, and the potential for a population-level shift in the distribution of NO_2 health effects.

US EPA reported heterogeneity in the responses of asthmatics exposed to NO_2. This variation may reflect differences in individual subjects and exposure protocols (e.g., use of mouthpieces vs. chambers to administer exposures, evaluation of effects during rest vs. exercise, participation by obstructed vs. nonobstructed asthmatics, and varying use of medications by participants) as well as potential differences in how specific investigators administer protocols.[18] US EPA did not quantitatively evaluate the possible impact of any these variables on the observed responses, except for comparing results based on activity level in a limited fashion (based on the observation that responsiveness to NO_2 is often greater following rest than exercise).

Subsequent evaluations of the NO_2-controlled exposure studies in humans illustrate how more refined meta-analyses (assessing the strength of responses), as well as meta-regressions (formally assessing dose-response relationships), can enhance insights regarding quantitative NO_2 exposure-response relationships and specific factors influencing such associations.[19] Similar to US EPA, Goodman *et al.*[19] evaluated the effects of NO_2 exposure (at concentrations ranging from 0.1 to 0.6 ppm) on AHR to airway challenges in a total of 570 asthmatics in 28 controlled exposure studies. Including studies of both specific and nonspecific responsiveness, this meta-analysis stratified analyses by factors demonstrated to affect AHR, i.e., airway challenge (specific/nonspecific), exposure method (mouthpiece/whole chamber), and activity during exposure (rest/exercise).[23,24]

The primary difference between the US EPA[18] and Goodman *et al.*[19] analyses is that while US EPA[18] evaluated only the percent of people with decreased AHR, Goodman *et al.*[19] also evaluated the magnitude of the change in AHR following NO_2 (vs. filtered air) exposure. They assessed magnitude by evaluating measurements of (1) the provocative dose of a challenge agent necessary to cause a specified change in lung function and (2) the change in FEV_1 after an airway challenge (e.g., Fig. 2). Goodman *et al.*[19] concluded that although several effect estimates from the meta-analyses were statistically significant, they were very small and not likely to be clinically relevant based on US EPA criteria defining moderate or severe outcomes.[25,26] Using meta-regression and effect assessments stratified by exposure level (0.1–0.2, 0.2–0.3 ppm, etc.), exposure-response associations were not statistically significant in the overall or stratified analyses (e.g., Fig. 3).

Compared to Goodman *et al.*,[19] US EPA[18] combined a greater number of studies in each meta-analysis because it did not use specific outcome measurements (e.g., the magnitude of the change in a specific lung function measure) but, rather, transformed each outcome to a binary variable (i.e., an increase or decrease in AHR based on any one of several possible measures). However, using this approach, US EPA[18] could not evaluate the magnitude of any effect or whether the effect increased in magnitude as exposure increased (i.e., exposure response). As a result, US EPA[18] was unable to fully assess whether observed statistically significant associations are causal, or whether effects were more indicative of homeostasis or adversity.

Because the NAAQS are intended to protect against adverse effects, the types of insights into causation and effect adversity afforded by the refined meta-analysis and meta-regression approach clearly could play an important role in determining protective standards. Moreover, as illustrated above, greater use of these approaches would provide greater insights into the quantitative nature of exposure-response relationships, as well as the key factors influencing such relationships. As a result, these powerful tools merit greater use in this research area than has occurred to date. It is noted that associations or effects are often identified as a basis for scientific and regulatory concern simply due to statistical significance; however, it is critical that such findings be interpreted within the full context of information needed to determine causality and adversity (e.g., Goodman *et al.*[27]).

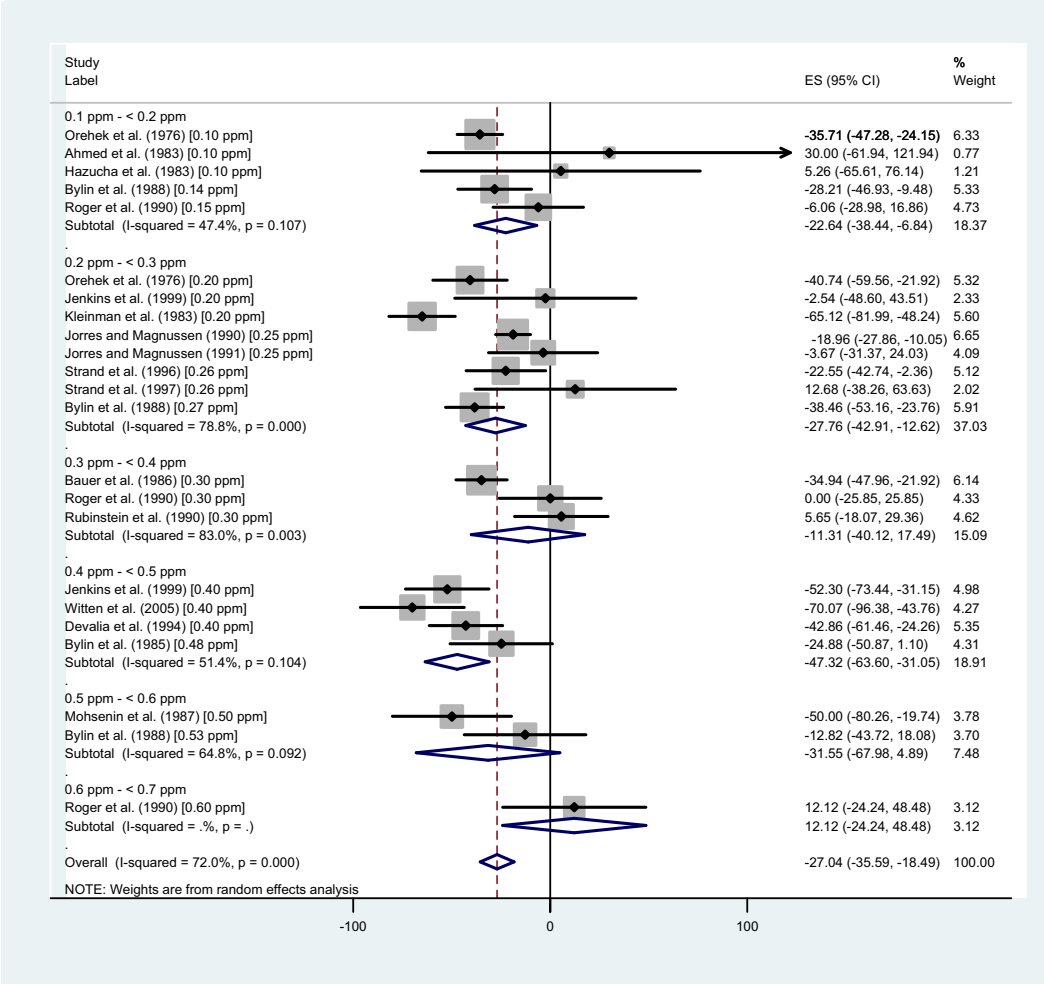

Fig. 2. Forest plot showing the difference in responses to airway challenge provocative doses following exposure of asthmatics to NO_2 vs. air. This figure illustrates the types of meta-analysis findings that can be graphically illustrated in a forest plot, e.g., the average change per dose from each study (the central dots within the squares), the proportional weights used in each meta-analysis (the squares), and summary measures and confidence intervals for each dose level (horizontal lines) and the overall study (the center lines and lateral tips of the diamonds). The results from individual studies and study combinations can be compared with the vertical lines (with the solid line indicating no effect and the dotted line indicating the overall summary measure) to assess whether results are consistent across studies or whether a dose-response relationship appears to exist. Adapted from Goodman *et al.*[19]

Meta-analyses of controlled exposure studies have not yet played a role in NAAQS evaluations of other criteria air pollutants. For O_3, several controlled human exposure studies have evaluated associations between exposure and adverse effects on lung function.[28–31] The majority of these studies reported no statistically significant changes after 6.6-hour exposures (with moderate exercise) to up to 0.06 ppm O_3. Because these studies are fairly homogeneous, a meta-analysis potentially could yield increased power to detect statistically significant effects. Unfortunately, the publications from these studies do not provide sufficient data to explore this possibility.

In these studies, healthy young adults were exposed to O_3 while exercising for up to 6.6 hours. Lung function was measured at several time points; however, complete data for all time points is not included in all publications of these studies. Omitting the data from the intermediate time points complicates data interpretation (e.g., by preventing researchers from evaluating whether findings reflect false positive results, a possibility that increases with an increased number of statistical comparisons). Similarly, in the

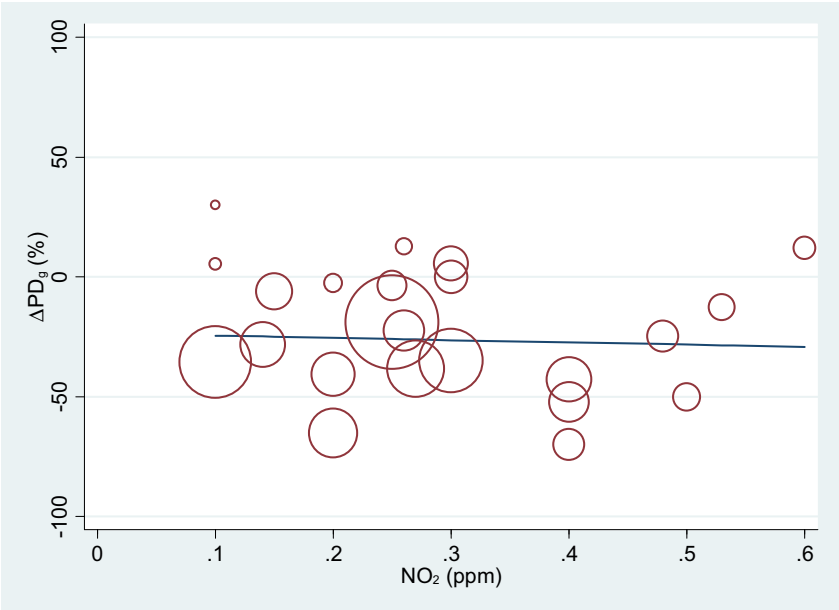

Fig. 3. Association between NO_2 exposure and airway hyper-responsiveness in asthmatics based on meta-regression of the difference between airway challenge provocative dose following exposure to NO_2 vs. air. This figure illustrates the use of a bubble plot to display meta-regression results. Each circle represents the findings from one study at a given exposure, while the area of each circle is proportional to the weight given to each measure in the meta-regression. Adapted from Goodman *et al.*[19]

absence of complete data, scientists conducting meta-analyses face greater challenges in assessing whether a statistically significant result is causal or simply a result of unaccounted for multiple comparisons and/or selection bias in the underlying studies. More complete data would also allow researchers to better evaluate other aspects of each study (e.g., the effects of exercise on the study observations).

Overall, the more focused nature of study conditions in controlled exposure studies enhances researchers' ability to design studies with greater consistency, making meta-analysis a particularly attractive tool for synthesizing findings from such studies. However, as illustrated in the examples presented above, reporting choices in the original study and the meta-analysis design can affect interpretation of findings. For example, if individual studies provide only a subset of the available data, subsequent data syntheses will be hampered and may yield biased results. Moreover, even when more complete information is available, meta-analyses of a specific topic can be interpreted differently, depending on how they are conducted. Important factors influencing interpretation include which studies are included in the analysis, the overall size of the data set, the specific exposure conditions evaluated, how outcomes are measured, how an outcome is considered (e.g., whether measurements are transformed), the approaches used to assess how various factors influence specific outcome measures, and how exposure-response analyses are conducted (e.g., using meta-regression).

Because controlled exposure studies play such an important role in US EPA's NAAQS determinations, using appropriate and rigorous methodologies to analyze such data is critical. Even in a relatively homogeneous class of studies such as controlled exposure studies, significant differences can exist among studies in the data that are collected, analyzed, and reported. Such differences can affect choices made in designing and implementing meta-analyses, and those choices have consequences for interpreting results.

3. OBSERVATIONAL EPIDEMIOLOGY STUDIES

Observational epidemiology studies explore the relationships between exposures and health outcomes in various populations, including the general population and population groups within specific exposure settings (e.g., workplaces). In contrast to controlled exposure studies, potential associations between exposures and outcomes are evaluated in "real-world" settings in observational epidemiology studies. As a result, researchers have far less control over study conditions and a greater degree of heterogeneity arising from aspects of study features and data analysis is inherent both within and among these studies (e.g., in population demographics and health status, types and nature of participants' exposures, measures of exposures and effects, and types and extent of confounding). Synthesizing findings across observational studies is also complicated by the

frequent lack of standardized approaches for presenting study methods and results, as well as the increasingly complex statistical methods used to analyze such data. While this inherent heterogeneity presents challenges in synthesizing study results, it also highlights the importance of applying tools such as meta-analysis to better understand and quantify, where possible, the bases for observed differences in findings (e.g., Stroup *et al.*[7]). In addition, when data from observational studies are used to support policy decisions, their heterogeneous nature presents more choices in selecting specific research areas warranting synthesis, more need to determine areas where synthesizing information will most effectively impact policy decision making, and more potential benefit in enhancing understanding of the available literature.

One heterogeneous element of observational epidemiology studies is the range of possible study designs, including time-series, cross-sectional, cohort, case-control, case-crossover, and panel studies. Researchers have used these study designs to assess both acute and chronic health effects associated with criteria air pollutants across a range of exposure durations. For example, standard cross-sectional studies examine exposure and outcome measures reflecting a single point in time, time-series studies examine exposure-response associations at multiple points over short time periods (e.g., days), and cohort studies typically follow study populations over long time periods (e.g., years or decades). In many cases, the studies are "opportunistic"; i.e., they are designed around data sources collected for other purposes, such as air data from fixed-site air monitors used to assess regulatory compliance. Exposure measures can also vary, e.g., the averaging time used to calculate air concentrations or the lag time between exposure and response measurement.

Ecological time-series studies are often used to assess health effects of short-term exposures to air pollutants. These studies compare daily population-averaged exposure estimates with daily population-averaged health endpoint counts (e.g., hospital admissions, emergency department visits, disease incidence or prevalence, and mortality). The relative rate of the endpoint (e.g., percent increase in mortality per unit increase in daily air pollution) is often calculated using either generalized additive models (GAMs) or generalized linear models (GLMs), two statistical modeling approaches that differ in their degree of flexibility and how confounding factors can be addressed.[32] Effect estimates can differ depending on the methodology used. Cohort studies, such as the American Cancer Society (ACS) Cancer Prevention Study, are commonly employed to assess health effects from long-term exposures.[33,34] In these studies, health effects are inferred based on differences in pollution levels between cities, rather than day-to-day differences in pollution levels in a single city. Because any factor that varies from city to city could be a potential confounder (including socioeconomic and lifestyle factors), controlling for confounding is particularly challenging.

For lead, many observational epidemiology studies have focused on effects associated with chronic exposures. Many such studies are cross-sectional in design, examining exposure and outcome data from individual studies or other sources (e.g., the National Health and Nutrition Examination Survey (NHANES)). Several cohort studies have also been conducted, including a set of studies initiated in the early 1980s that were coordinated to a limited degree to help researchers compare and synthesize results.[12,35]

For a number of criteria air pollutants (e.g., PM, SO_2, and NO_2), US EPA has focused on respiratory effects, cardiovascular effects, reproductive and developmental effects, cancer, and mortality. Within these general categories, US EPA has evaluated many specific endpoints. For example, for respiratory effects associated with $PM_{2.5}$ (particulate matter <2.5 μm in diameter), studies have evaluated asthma, pulmonary function, respiratory symptoms, hospital admissions, and emergency department visits, as well as markers of pulmonary inflammation or injury.[14] In the case of lead, US EPA has assessed its potential causal role in 25 types of health effects in seven categories of organ systems or effects.[12]

Approaches used to estimate exposure represent an important source of heterogeneity. For example, because the composition and particle size distribution of different types of PM vary, studies can focus on specific types of PM (e.g., diesel exhaust particles), PM source areas (e.g., urban or rural settings), or specific size fractions (e.g., total suspended PM, particulate matter <10 μm (PM_{10}), or $PM_{2.5}$). Because these different types of PM may not be comparable, it is not always appropriate to combine studies that evaluate them. To address this problem, some researchers have applied conversion factors (e.g., to convert PM_{10} to $PM_{2.5}$[36]); however, this approach can introduce uncertainty into data analyses because such factors can vary under different site-specific conditions.[37]

Heterogeneity can also exist in approaches to measuring exposure (e.g., using personal, ambient, or indoor air measurements) or averaging exposure duration; e.g., health effects of O_3 exposures have been evaluated based on a daily average, a daily maximum eight-hour average, or a daily one-hour maximum. Results based on different exposure estimates cannot easily be combined. To address this issue in meta-analyses, some researchers have used conversion factors to derive estimates reflecting a uniform averaging time.[38] These conversion factors, however, have been shown to introduce error and distort observed pollution patterns, and can result in biased health effect estimates.[39]

Among the criteria pollutants, estimating exposure is particularly challenging for lead. As a result of historical lead uses and environmental distribution, lead exposures can occur via numerous environmental media in addition to air (e.g., soil or drinking water), and historical as well as stored sources (e.g., bone lead) from past exposure are relevant for exposure assessment. In addition, exposure modeling for lead (e.g., to assess potential impacts of air emissions on human exposures) is a multistep and multifaceted process. Moreover, because most epidemiological studies assess lead exposure levels based on biomonitoring data (e.g., blood lead levels), which reflect an integrated measure of lead exposures across media and time frames, studies reflecting exposures from a range of sources (e.g., including dietary or drinking water sources) are relevant in health evaluations for lead. Thus, the exposure characterization for lead encompasses a diverse spectrum of potential sources, measurement methods, and modeling approaches.

3.1. Meta-Analysis Applications

Meta-analyses and related data synthesis methods (e.g., pooled analyses) have played an increasingly important role in NAAQS evaluations of observational epidemiology studies. Such analyses include traditional meta-analyses, e.g., where published results from air pollution studies conducted in individual cities are combined using meta-analytical methods to obtain a summary estimate. Air pollution research also includes multicity studies (a type of pooled analysis), where city-specific effects are estimated using a common analytic framework. Applying similar techniques to those used in meta-analysis, these effects are then combined to estimate effects across all cities.[5] The use of consistent analytical approaches in multicity studies helps to reduce the heterogeneity commonly encountered when combining results from more disparate individual studies. Moreover, considering all of the raw data collected in a multicity study (rather than focusing only on published results) avoids publication bias. These approaches are particularly useful in exploring sources and impacts of heterogeneity, increasing statistical power to detect effects, yielding more broadly applicable overall effect estimates, and evaluating concentration-response relationships.

The largest U.S. multicity time-series study is the National Morbidity, Mortality, and Air Pollution Study (NMMAPS).[40] Using data collected in 90 cities, researchers examined associations between short-term exposures to air pollutants (including PM_{10}, $PM_{2.5}$, and O_3) and mortality.[41–44] Researchers have explored the study findings regarding health impacts of air pollution on a city, regional, and national level (e.g., Fig. 4). Specifically, semi-parametric regression statistical models (e.g., GAMs or GLMs) have been used to estimate city-specific effects; researchers have then applied hierarchical statistical models to estimate national effects, develop concentration-response functions, and evaluate sources of heterogeneity across cities.

In addition to the multicity studies, several reviews and meta-analyses of air pollution epidemiology studies have been conducted, primarily of time-series mortality studies.[32,36,38,45–48] Some of these studies evaluated the health impacts of several air pollutants (e.g., PM_{10}, SO_2, NO_2, CO, and O_3), both independently and in a multipollutant context.[36] Others have focused on only one pollutant (e.g., O_3), but considered potential confounding by copollutants (e.g., PM).[38,48] Overall, these reviews have explored the role of confounding factors in exposure-outcome relationships, identified systematic approaches for applying time-series studies in meta-analyses, and evaluated the relative merits of various statistical models and how well they estimate health effects.

For the past 20 years, researchers have used meta-analysis to synthesize diverse aspects of the lead literature, including studies addressing health

[5] In the multicity studies, the hierarchical data set includes subjects within cities on the first level and cities in the second level. These types of multilevel analyses allow for using study characteristics as potential explanatory variables that can shed light on differences in outcomes across cities.

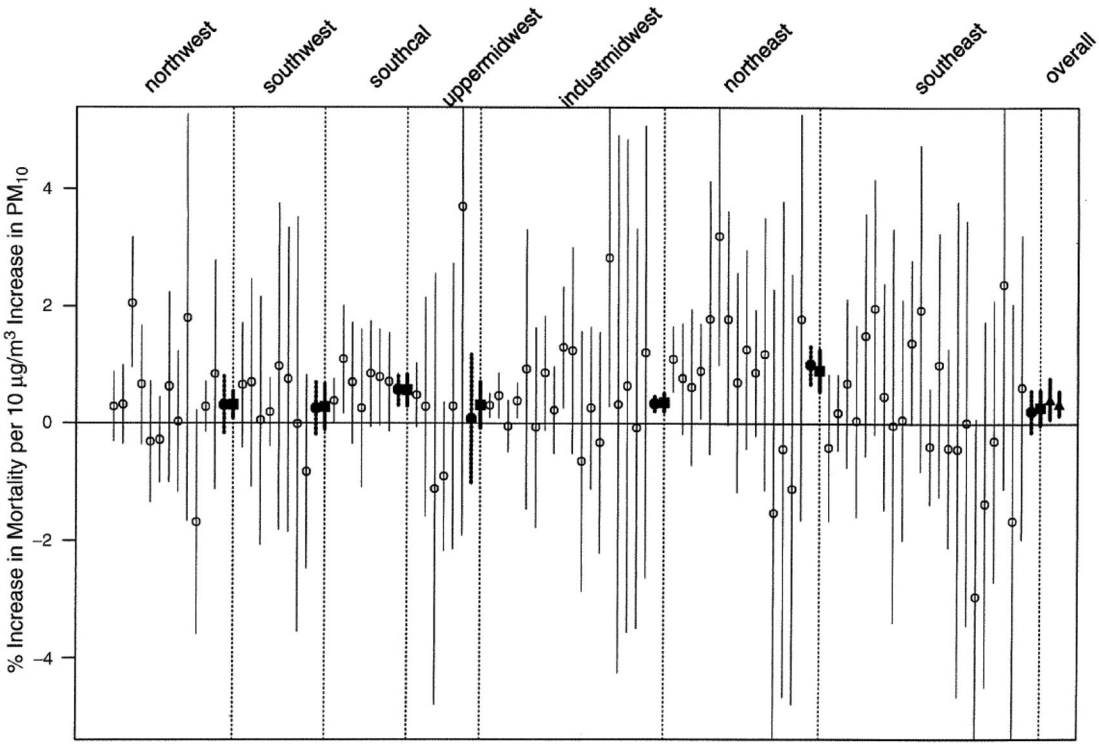

Fig. 4. Percent increase in total mortality associated with a 10 $\mu g/m^3$ increase in PM_{10} in 90 NMMAPS cities, with 95% confidence intervals and grouped by region. This figure illustrates another approach to displaying results using a forest plot. The open circles represent specific cities. Summary estimates based on two methodologies are shown, in bold, to the right of the individual city results for each region (as delineated by the dotted lines) and for national estimates (shown on the far right). Adapted from Samet *et al.*[40]

effects (e.g., neurocognitive effects;[49] behavioral effects;[50] cardiovascular effects such as blood pressure impacts;[51] and cancer[52]), potential effect markers (e.g., genetic polymorphisms for δ-aminolevulinic acid dehydratase, an enzyme involved in heme biosynthesis (ALAD)[53]), and exposure issues such as effectiveness of interventions to reduce exposure.[54] These analyses have estimated associations between lead exposure and various outcomes, odds or risk ratios, and concentration- or dose-response relationships.

Overall, meta-analyses (and multicity studies) of observational epidemiology studies are but one component of US EPA's descriptive discussions of specific health endpoints in its NAAQS evaluations. For example, for lead, meta-analyses have been included in discussions of blood pressure[51] and conduct disorder.[55] For several other criteria air pollutants (e.g., PM and O_3), multicity studies have played a fundamental role in policy decisions regarding the NAAQS. These meta-analysis applications provide numerous opportunities to identify ways that this technique could be used more systematically and effectively in the future.

3.2. Meta-Analysis Interpretation Issues

Like the meta-analyses conducted using NO_2 controlled exposure studies, the observational epidemiology literature regarding criteria air pollutants illustrates how differences in meta-analysis choices can affect results and interpretations. As illustrated below, two key issues to consider when interpreting meta-analyses findings are bias and heterogeneity.

3.2.1. Bias

Meta-analysis applications in air pollutant research reflect the importance of carefully considering potential bias in study data collection, presentation, analysis, and synthesis. In particular, any systematic differences in the analytical approaches used in individual studies can result in bias in a meta-analysis. As one example, the meta-analyses of O_3 time-series

studies illustrate how bias can occur based on the averaging time or lag period between exposure and effect chosen when analyzing and reporting the study results (i.e., selection bias). For example, both Levy *et al.*[48] and Bell *et al.*[38] found higher effect estimates based on results reflecting the most statistically significant lag times reported in individual time-series studies versus using a consistent lag time of 0 in the effects analyses. This type of bias can be avoided by analyzing and reporting the data using a single consistent averaging time or lag time for all studies. In a meta-analysis of asthma incidence and long-term air pollution exposures, Anderson *et al.*[56] reported publication bias, as well as bias associated with unsystematic data analyses (i.e., using only effect estimates that were statistically significant rather than selecting estimates based on *a priori* selection criteria). Similarly, bias in synthesized results can also arise due to differences in approaches taken to address seasonal trends in health effects of interest (e.g., mortality and asthma) that are unrelated to O_3 exposures. For example, in time-series analyses, researchers often use a spline-based nonlinear function to represent the trend of a health effect over time, with the function's degrees of freedom describing the trend. Some studies use a fixed degrees of freedom based on biological knowledge or previous work regarding trends (e.g., an observation that the asthma attack rate is greater during the spring); while others may use an estimated degrees of freedom based on the observed data. Because multicity studies are specifically designed to use consistent study designs and data analysis approaches, these aspects of bias are minimized in these studies.

Meta-analyses of observational air pollutant epidemiology studies also indicate a potential source of bias in results from single-city studies relative to coordinated multicity studies. For example, as shown in Fig. 5, Bell *et al.*[38] observed that the mortality estimates from their meta-analyses of single-city studies were significantly higher than those from the NMMAPS summary multicity estimates. The authors suggested publication bias as the basis for this finding, i.e., the NMMAPS multicity reports may more routinely include negative results from individual cities, while researchers conducting single-city studies may be less likely to submit negative findings for publication.

Impacts of bias on study interpretation also arise in lead studies. For example, Bellinger[57] identified numerous potential opportunities for researchers to selectively report or introduce bias in neurotoxicity studies based on their choices regarding which

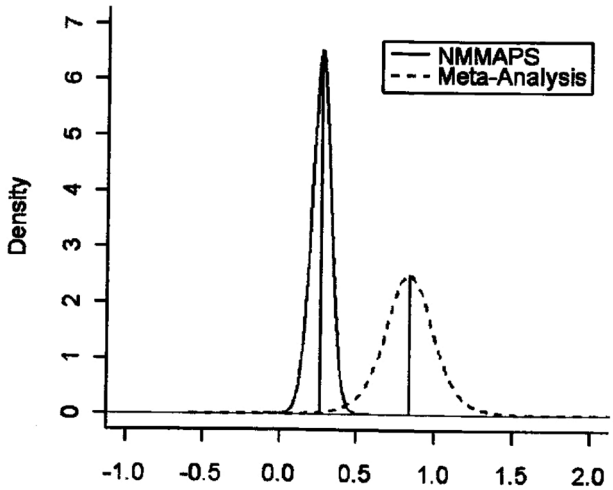

Fig. 5. Distributions of summary log-relative risks of all-cause mortality associated with a 10-ppb increase in O_3 in 95 cities (NMMAPS) compared to a meta-analysis of 11 U.S. estimates. The multicity results yielded a lower and more precise estimate of the overall percent decrease in mortality associated with O_3 exposures than did the meta-analysis based on studies reporting results from single cities. Source: Bell *et al.*[38]

confounders to adjust for and how to parameterize them, how to parameterize exposure (e.g., quartiles vs. quintiles, or linear vs. log-transformed), how to express exposure/outcome associations (e.g., highest vs. lowest quintile or piece-wise regression slopes), and which analyses "among the myriad typically conducted" to present. Bellinger[57] also illustrated how conclusions can vary depending on choices made in addressing covariates in the data analyses. In one instance, conclusions regarding potential associations between lead exposures and the results of a continuous performance test (a measure of attention and neurological functioning) depended on the statistical criteria (i.e., p values) used to select covariates included in the analyses.[58]

In another example, a series of communications regarding a pair of meta-analyses of associations between neurobehavioral effects and occupational lead exposures highlights the need for clear transparency in how data are selected and extracted to understand any sources of bias.[59–64] Specifically, one meta-analysis of 22 studies of neurobehavioral effects in occupational populations exposed to lead (with blood lead concentrations less than 70 μg/dL) concluded: "The data available to date are inconsistent and are unable to provide adequate information on the neurobehavioral effects of exposure to moderate blood concentrations of lead." [61] Seeber *et al.*[63] determined that the "conclusions from

published results about neurotoxic effects of inorganic lead exposures <700 μg lead/L blood [70 μg/dL] are contradictory at present"; however, the authors also noted that available test results "provide evidence for subtle deficits being associated with average blood lead levels between 370 and 520 μg/L [37 and 52 μg/dL]." These research groups debated a number of choices made in compiling and analyzing the component studies, including whether study quality and potential confounding factors were adequately accounted for, whether the number of neurobehavioral test measures showing significant results (reported to be two out of 22 measures in Goodman *et al.*[61]) was appropriately addressed, and the degree to which performance prior to lead exposure was accounted for in the underlying studies.

A study of attention deficit hyperactivity disorder (ADHD) symptoms in young children (i.e., inattention and hyperactivity/impulsivity) illustrates how meta-analyses cannot remediate fundamental limitations in the underlying studies. Specifically, a meta-analysis estimated the average ADHD-related effect size in 33 studies of children and adolescents (i.e., associations between ADHD symptoms or diagnoses and various measures of lead exposure.[50] As recognized by US EPA's Clean Air Science Advisory Committee in its review of US EPA's *Integrated Science Assessment for Lead*,[65] most studies of potential associations between lead exposure and ADHD fail to include or adequately assess information regarding parental psychopathology; this fundamental factor may play an important role in children's ADHD status "via parenting behavior, and/or genetic contributions to disorder type." As acknowledged by Goodlad *et al.*,[50] "the conclusions that can be drawn from the current study are limited by the methodological designs of the studies that were analyzed" (including the lack of information regarding parental ADHD status) and "these studies and the meta-analysis of these studies describe the association between lead burden and ADHD symptoms and cannot be used to draw strong causal conclusions." Clearly, an essential component of any meta-analysis is a sound understanding and acknowledgment of any limitations or other notable features of the included studies.

3.2.2. Heterogeneity

Heterogeneity is a pervasive challenge in synthesizing results from observational epidemiology studies. As a result, air pollutant researchers have worked to make certain categories of available studies more consistent and make it easier to more effectively combine and compare study results. As described above, as one means of enhancing data integration, some researchers have conducted multicity studies applying a common research approach at numerous sites.

Despite efforts to reduce the effects of heterogeneity on air pollution research, studies continue to differ with respect to such factors as outcome definitions, study populations (e.g., age), study periods (e.g., seasonal vs. year-round analyses), and statistical methods (including approaches to assessing confounding by co-pollutants and other factors). For example, in time-series studies of air pollution and mortality, a principal issue is how confounding by temporal cycles and weather is addressed. In one instance, researchers identified software issues related to the use of the GAM model that suggested this model overestimated effects for some air pollutants; these issues were subsequently corrected.[46] For O_3, issues also arise because U.S. air monitoring data are only collected in the summer in many areas. In addition, researchers have used different exposure averaging times to evaluate O_3 effects (e.g., 24-hour average, 8-hour average, and 1-hour maximum), yielding mortality estimates that are not equivalent. If these differences are not addressed appropriately in data analyses (including meta-analyses), unreliable conclusions may be drawn.

The lead literature also reflects examples of efforts undertaken to reduce the heterogeneity inherent in observational epidemiology studies, as well as the impacts of such efforts on lead health effects research and regulatory applications. As noted above, one such effort began in the early 1980s, when researchers in the United States and several other countries undertook a coordinated set of prospective cohort studies using similar research protocols (hereafter, "the longitudinal lead studies"). Focusing primarily on the neurocognitive development of participants (e.g., as reflected in IQ measures), evaluation of study subjects began prior to birth and has extended, in some cases, to young adulthood.[66] The comparability of the longitudinal lead studies' designs has fostered numerous publications over the past three decades, including comparative discussions within specific study reports and comparative evaluations of such issues as the age range thought to represent the most susceptible period to lead effects.[67] As discussed below, efforts to enhance consistency among the longitudinal lead

studies have not removed all barriers to effective data synthesis.

While the researchers involved in the longitudinal lead studies worked to enhance the consistency of certain elements of study design and implementation, an early effort to conduct a meta-analysis based on 35 reports from five of these studies observed that the approaches used to analyze and report the study results were less consistent. In particular, Thacker *et al.*[68] found that it was not possible to compile the data from these studies because they differed regarding the statistical approaches that were used to summarize the study observations (e.g., data transformations, such as treatment of blood lead data as a categorical or continuous variable, and statistical summary parameters, such as regression coefficients, correlations, and changes in standardized scores) and they provided insufficient information to develop a consistent set of statistical measures. More fundamentally, few overlaps were observed in the times at which blood lead concentrations and IQ were measured in the studies. Other factors hampering meta-analyses included conflicting results and inconsistent patterns of regression and correlation coefficients (i.e., heterogeneity[68]). As a result, despite efforts to enhance study comparability, they were insufficient to support the more detailed comparisons and analyses of a formal meta-analysis. To support greater consistency in study reporting and data collection in a centralized location, Thacker *et al.*[68] urged development of a registry for the longitudinal lead studies.

Subsequent efforts using pooled data (not summary estimates) from a subset of these studies have played a central role in US EPA's efforts to set air standards for lead as well as other regulatory and risk assessment settings.[69] However, this pooled analysis reflects only a small portion of the health effects literature available for lead. Moreover, researchers have noted that studies of neurodevelopmental impacts (of lead and other substances), as well as other areas of epidemiological research, would benefit if researchers used more consistent analytical and reporting approaches.[57,70,71] In particular, focusing on neurotoxicity data, Bellinger[57] advocated for the development of "consensus standards for the conduct, analysis, and reporting of epidemiologic research . . . [to] enhance the credibility of the data generated (and of the field as a whole), as well as the ease with which the results of different studies can be compared and combined in meta-analyses."

3.3. Future Directions

The types of challenges discussed above for synthesizing data from observational epidemiology studies are not limited to air pollutant studies, and other attempts at meta-analysis have led to similar conclusions regarding the need to improve data collection to better support data synthesis. For example, despite identifying approximately 40 publications addressing studies of 11 cohorts, researchers exploring the possibility of conducting a meta-analysis regarding associations between neurotoxicity and polychlorinated biphenyl (PCB) compounds concluded that the "studies were too dissimilar to allow a meaningful quantitative examination of outcomes across cohorts."[72] They note that studies of neurodevelopmental toxicity might be particularly vulnerable to heterogeneity due to the large number of test batteries available (often with numerous combinations of subtests) and varying options for scales and cutoff points for categorizing results. To better support meta-analysis efforts and study comparisons, these researchers recommend that future research efforts continue to use assessment measures and exposure assessment methods that are comparable to previous methods, even as new methods are developed. They also recommend development of specific assessment and statistical methods to be used in studies, as well as approaches for greater data sharing (e.g., as a component of research funding requirements) and data/analysis archiving (e.g., by journals). Their general recommendations are clearly transferable to other research areas. Moreover, as recognized by Meyer-Baron *et al.*,[73] enhancing efforts to systematically synthesize and summarize available research findings is not only important for developing sound study interpretations, but is also increasingly important for more effectively identifying research gaps that most warrant use of decreasing research funding resources.

Another challenge in air pollution research is how to correctly assess the effects of individual air pollutants and evaluate confounding effects and other interactions of co-pollutants and other factors, such as nonchemical stressors (e.g., socioeconomic variables).[74] Although some efforts have been made to meta-analyze multipollutant data, the lack of consistently reported results from multipollutant analyses has hindered proper data synthesis.[36] Opportunities remain for better data evaluation and reporting to enhance synthesis across such studies.

New statistical techniques are advancing and improving the use of observational epidemiology data in meta-analyses. For example, Bayesian hierarchical statistical techniques, which are being implemented in multicity study analysis, provide opportunities for evaluating factors that contribute to heterogeneity in single-city studies, including the potential for real differences in effects to exist among cities. In such analyses, a two-stage Bayesian hierarchical model can be developed in which the effects observed in each city are assumed to differ but to follow the same distribution (with a distribution mean of the overall multicity effect and a distribution variance quantifying the heterogeneity among cities). Hybrid meta-analytic approaches are also being developed to incorporate uncertainty associated with combining information from a limited number of studies.[74,75]

Overall, the diverse air pollutant observational epidemiology literature presents many opportunities for applying meta-analysis approaches and learning how to refine and improve such approaches. As illustrated above, approaches used to report study data can influence whether researchers can synthesize study findings, or whether bias is introduced into analyses (e.g., where researchers selectively report only a subset of study findings). Most notably, the extensive inherent heterogeneity has spurred researchers to develop approaches for encouraging greater study consistency in certain research areas (e.g., implementing multicity studies of air pollution exposures or encouraging development of guidelines for neurotoxicity studies). As revealed by previous efforts to implement more consistent research approaches, such efforts must take a broad perspective on the concept of study consistency. In particular, to enhance the ability of researchers to synthesize results from multiple individual studies, consistency guidelines should consider issues associated with data analysis and reporting, as well as study design and implementation.

4. TOXICITY STUDIES

To date, NAAQS levels and averaging times have been based primarily on human data, but the underlying causation evaluations also consider toxicology data. Animal studies often examine endpoints relevant to the causal questions posed by the NAAQS process. Although meta-analyses have traditionally been used mostly for human data,[76] they can be helpful for synthesizing animal data for specific endpoints and determining whether those data are robust.

Heterogeneity in animal toxicity studies can arise due to the use of different species, study designs, and protocols. However, they may be more homogeneous than observational epidemiology studies because researchers can better control exposures, test conditions, and outcome assessments. Thus, evaluating data from laboratory animal studies may help elucidate issues raised in epidemiology studies or in meta-analyses of those studies. Using meta-analysis to evaluate animal study results could also encourage researchers to use more consistent study designs that in turn would strengthen subsequent meta-analyses. In addition, the increased precision of meta-analyses as compared to analyses of individual studies can help reduce the number of laboratory animals used in research; a meta-analysis of existing data may prove to be a more effective and informative use of research resources than a new primary experiment in animals when none of the previous experiments asking the same biological question have had sufficient statistical power.[77]

Many types of data from experimental studies using laboratory animals can be summarized and quantified using meta-analysis approaches. Animal study data may be binary (e.g., pregnancy, mortality), categorical (e.g., low, medium, or high amount of cellular damage in a particular organ), or continuous (e.g., blood pressure, lung function decrements). The data may also be presented as counts or percentages, such as the total number or percentage of treated animals with a specific tumor type. The methods for analyzing these data can also vary. In a review of 46 published meta-analyses of laboratory animal studies, Peters *et al.*[77] determined that researchers most commonly used simple methods to quantitatively synthesize results across studies, such as calculating mean or median values of outcome measures. They have also applied other methods, such as fixed- and random-effects precision-weighted models and exposure-response models.

For example, Valberg and Crouch[78] conducted a meta-analysis of data from eight studies of lung tumors in rats following lifetime inhalation of diesel exhaust particulates (DEPs), evaluating statistical evidence of a threshold in lung tumor response between high and low exposure concentrations. They used a multistage model to determine maximum likelihood estimates and upper confidence limit estimates of the exposure-response slope, concluding that the tumor responses observed at high levels of DEP exposure do not occur at low exposures. By contrast, in a meta-analysis of organ toxicity in laboratory animals exposed to nano-titanium dioxide, Chang *et al.*[79] used

a simpler approach based on the number of studies with positive findings at each dose for each endpoint. The authors stated that because of the variety of animal species and endpoints included in the studies, it was not possible to calculate a summary estimate of effect size. They determined that the pattern of positive results for the *in vivo* toxicity of nano-titanium dioxide depended on the dose, exposure route, and organ examined. They also observed that the highest percentage of positive studies reported effects in the liver and kidney. These findings were not evident by reviewing the individual studies.

Meta-analyses of animal toxicity studies can help determine whether observed effects of chemical exposures are consistent and readily generalized, but several factors must be considered. As with human data meta-analyses, publication bias can significantly affect interpretation of animal data meta-analyses, yielding overestimates of treatment-related effects. In addition, as noted above, between-study heterogeneity is a common meta-analysis feature that must be addressed. Some heterogeneity arises because studies differ in the animal species used. However, studies using different species can be combined in a meta-analysis if there is evidence that the outcome of interest works by the same mechanism across species or if species differences are accounted for in the statistical models.[77]

A major problem associated with animal data meta-analyses is the large number of published studies that incompletely report study design and methods. No widely used guidelines exist for reporting results from individual animal experiments, so the quality of primary studies varies. High-quality studies with detailed experimental information will facilitate high-quality meta-analyses. Missing information for a given parameter can introduce bias into the study, as well as any meta-analysis incorporating the study. Failure to consider study differences in the statistical models as a result of missing information can also yield reduced statistical power and false positive results.[80] If possible, all experimental variables should be incorporated into the analysis. Adhering to high-quality standards for conducting and reporting experiments can reduce the confounding effects of bias and enhance the validity and precision of the results.

In recent years, several investigators have proposed guidelines for reporting laboratory animal data in primary studies to improve the quality of scientific publications and facilitate meta-analyses and systematic reviews.[77,81–84] For example, Hooijmans *et al.*[82] developed a "gold standard publication checklist" of items that should be included in every published animal study and Kilkenny *et al.*[83] recommend the use of ARRIVE (Animals in Research: Reporting *In Vivo* Experiments) guidelines, a 20-item checklist describing the minimum information that all scientific publications reporting animal research should include. In addition to general information on the study design and methods, each set of guidelines recommends a sample size calculation prior to starting the study. In a related effort to strengthen animal studies and their usefulness, a recent review focused on methods for assessing the risk of bias, identifying 30 approaches that have been used (including approaches applied in some of the guideline documents discussed above).[85]

Although they are not currently used in causation evaluations supporting NAAQS determinations, meta-analyses of animal toxicity studies can improve interpretation of existing results from primary studies, which can inform causality determinations by providing plausibility for associations observed in human studies. Meta-analysis offers a framework for investigating potential publication bias, which can lead to overestimates of treatment effects and make the evidence unreliable for regulatory decision making. If the sources of bias in laboratory animal studies are better understood, these studies may be conducted and reported using higher quality approaches. Such improvements in the underlying scientific studies would help regulatory decisions to be made based on high-quality, unbiased data.

5. MECHANISTIC STUDIES

As with studies of animal toxicity data, studies reporting mechanistic data that are considered in the causality evaluations for the NAAQS can also be amenable to meta-analysis. Combining data from multiple mechanistic studies in a meta-analysis can help scientists better understand the mode of action (MoA) of a particular chemical and the biological plausibility of health effects reported in studies of humans or animals.

There are many types of *in vitro* mechanistic data that can assist understanding of the toxicity of chemicals at the cellular or molecular level. These types include data regarding cytotoxicity, enzyme activities, apoptosis, inflammation, cell proliferation, genotoxicity, cell transformation, genetic polymorphisms, and expression of genes, proteins, or

metabolites. Similar to animal data, mechanistic data can be binary, categorical, continuous, or reported as counts or percentages, and these data can be combined across studies using multiple statistical methods. Between-study heterogeneity is also an issue with mechanistic studies, as cell types and tissues from different species—maintained under different *in vitro* conditions and subjected to different protocols—can be used to explore the same biological question.

One category of mechanistic study that provides a good opportunity for meta-analysis is global gene expression studies using microarray technology. Evolving over the past two decades, this technology is being used in a wide array of contexts, providing an opportunity for innovative meta-analysis applications. Although these studies can generate a large amount of data that require reliable interpretation, they often have a relatively small sample size, as the simultaneous expression of tens of thousands of gene probes is typically examined in only tens or hundreds of biological samples. Combining gene expression studies through meta-analysis yields a larger data set, which increases the statistical power to more precisely estimate treatment- or exposure-related differences in gene expression. The increasing public availability of raw data from microarrays in various repositories greatly enhances the feasibility of conducting a meta-analysis of gene expression studies.[86] There are many gene expression meta-analyses in the published literature, and Ramasamy *et al.*[87] outlined practical guidelines for conducting a meta-analysis of microarray data sets in seven distinct steps.

Several challenges exist for conducting meta-analyses of gene expression data. One challenge is the data quality regarding reporting of phenotypic information about the biological samples examined. A set of criteria called MIAME (Minimum Information About a Microarray Experiment) was developed to guide researchers in providing information on the necessary experimental conditions for verifying and reproducing microarray study results.[88] Microarray data submitted to public repositories, as well as to many scientific journals for publication, must be MIAME-compliant, but information often is incomplete regarding the biological properties of samples and the phenotypes that were assayed, including the sex and age of the organism or tumor information (e.g., stage, grade, metastasis) for cancer studies.[89] To ensure the reliability and overall quality of meta-analyses of these studies, it is necessary to include as much biological information as possible when individual gene expression studies are reported.

Another challenge is that the results of a meta-analysis of gene expression studies can often be dominated by an outlying study, which can be a significant problem when analyzing thousands of genes simultaneously within the "noisy" environment of a microarray experiment. Outlying data can reduce the statistical power of the study, but methods that combine robust rank statistics can alleviate this issue.[80]

A further challenge for combining gene expression data from multiple studies is the technical complexity of integrating data across multiple microarray platforms. Many microarray platforms are available, with overlapping sets of gene probes across platforms. While some normalization procedures require all studies in a microarray data meta-analysis to use the same platform for merging data sets, some investigators have developed advanced normalization techniques to eliminate between-study heterogeneity due to varying platforms and allow data sets to be directly merged.[80]

Examining gene expression changes in cells or tissues from different species can also be a source of between-study heterogeneity, as large variability often exists between gene expression patterns from different organisms. Conversely, combining data sets from multiple species can increase the potential to detect gene expression changes related to biological processes that are evolutionarily conserved across species, which can support a hypothesized MoA. Statistical methods for reliable cross-species analyses of gene expression data have been proposed by several investigators (as reviewed by Kristiansson *et al.*[90]).

In addition to gene expression and other types of *in vitro* studies, *in vivo* mechanistic data from studies in laboratory animals (such as those discussed in the previous section) and humans can also be combined using meta-analysis methods. For example, Nakao *et al.*[91] investigated whether the behavioral and cognitive deficits of ADHD are associated with underlying structural and functional brain abnormalities in humans. Specifically, they combined data from 14 structural neuroimaging studies of gray matter abnormalities in the brains of ADHD patients and healthy control subjects and used meta-regression methods to examine the effects of age and use of stimulant medication on gray matter volume in specific brain areas. Similar investigations of brain structure and function have also been undertaken in lead-exposed individuals as one component of recent interest in ADHD by lead health effects researchers.[92–95] Like gene expression studies, structural neuroimaging studies are an example

of a relatively new research tool that is applied in an increasing range of contexts and generates large amounts of data; thus, such studies are well suited to being combined using meta-analysis techniques.

It is important to use all of the available information on a chemical at relevant doses or exposure levels when evaluating the likelihood that exposure can cause adverse health effects, including data from mechanistic studies. Although not yet commonly used for mechanistic data, meta-analysis can be an objective method for combining the results of these studies in causality determinations for NAAQS as more mechanistic studies are conducted. By helping to synthesize results regarding a chemical's MoA, meta-analyses of mechanistic studies can help researchers understand whether associations from epidemiology and animal toxicity studies are biologically plausible. For example, data from mechanistic and toxicity studies might suggest a certain MoA for toxicity; however, if the toxicity data indicate a threshold for effects that is much greater than the exposures associated with epidemiology study findings, synthesis of such findings would raise questions regarding the causal nature of the associations observed in the epidemiology studies. Such analyses can provide a more objective basis for regulatory decisions.

6. DISCUSSION

Meta-analysis provides a useful framework that offers many benefits for systematically organizing, synthesizing, and interpreting data for a wide range of research areas and study types. As illustrated in this article, meta-analysis is adaptable to many types of outcomes, study designs, and categories of outcome measures. Meta-analysis can be particularly useful for identifying and exploring the impacts and sources of heterogeneity in study results, a factor that is particularly prevalent in observational epidemiology studies. Such tools can also be useful for identifying limitations common to many studies and examining factors that may influence perspectives on overall study findings. Where suitable data are available, meta-analysis or meta-regression can help determine the overall magnitude of outcomes reflected in study findings. For research areas where individual study sizes are often relatively small (e.g., human controlled exposure and animal toxicology studies), data aggregation via meta-analysis can strengthen researchers' ability to draw well-supported conclusions.

Meta-analysis techniques can also allow researchers to draw upon a broader database when conducting analyses to support policy determinations. For example, when establishing an effects screening level for long-term exposures to nickel in air, the Texas Commission on Environmental Quality (TCEQ) derived a unit risk factor (URF) for potential carcinogenic effects using a meta-analysis approach.[96] Specifically, instead of deriving a toxicity value based on a dose or exposure level from a single study, TCEQ integrated three values from two studies of lung cancer in nickel-exposed workers to derive a final URF that reflected the relative value and significance of the data from each study. Moreover, where research or policy questions draw upon findings from a variety of areas and disciplines (e.g., as is required in causality determinations or WoE evaluations based on epidemiology, toxicology, and mechanistic studies), the results from sound meta-analyses of multiple individual components can be integrated to yield a stronger foundation for the ultimate question of interest.

Both the outcome of specific meta-analyses and the meta-analysis process itself can yield useful insights for answering scientific questions and making policy decisions. Most notably, conducting the systematic study review required for a meta-analysis can help researchers resolve and understand the basis for apparent inconsistencies among individual study results. Comparing studies in a meta-analysis framework can also help identify false positives, insights regarding factors influencing study results, and whether study findings can be generalized to other populations. Furthermore, sensitivity analysis within a meta-analysis framework can indicate how robust available data are and how specific study results influence the overall study findings. Even where data are insufficient or unsuitable for conducting a meta-analysis, the process of reviewing the available data within a meta-analysis framework can help researchers identify important factors influencing the study outcomes or critical data gaps that need to be explored in future research.

Despite its many strengths, meta-analysis cannot resolve all data interpretation issues. In particular, meta-analyses cannot yield insights regarding missing data elements or resolve limitations in the underlying data (e.g., inadequately addressed potential confounding factors). Meta-analysis tools cannot be applied in all circumstances—and cannot directly encompass all available data regarding a research

question; the studies included in a meta-analysis must all address the same research question in the same way (e.g., using the same endpoint measures). As noted in one of the earliest sets of guidelines for conducting meta-analyses of environmental epidemiology studies,[4] meta-analysis may not be useful when the relationship between the exposure and outcome is obvious, only a few studies are available for a particular exposure/outcome relationship, access to data of sufficient quality is limited, or substantial variation exists in study designs or populations. In addition, important differences in effect estimates, exposure metrics, or other factors may limit or even preclude quantitative statistical combination of multiple studies. Conversely, studies excluded from a meta-analysis may form a critical part of the context for interpreting the meta-analysis results, e.g., by providing useful information to be included when qualitatively discussing the results.[4]

Moreover, meta-analyses alone cannot address the adversity of the studied outcome. Instead, such determinations must consider the degree to which the outcome measure is related to the actual adverse effect of concern or reflects functional impairment. For example, Goodman *et al.*[27] reported statistically significant effects associated with short-term exposures to SO_2 concentrations in controlled exposures studies that were transient, reversible, and of low severity, and concluded that the effects were not likely to be adverse. Because meta-analyses have more statistical power to detect associations than individual studies, statistically significant associations that do not necessarily reflect an adverse effect are more likely to be reported. This example also demonstrates that other information must be brought to bear to determine the toxicological or clinical significance of statistically significant study results.

The examples discussed in this article illustrate other key issues for meta-analysis and opportunities for using meta-analysis in air pollutant research and policy applications. For example, analyses of controlled exposure data for NO_2 demonstrate how design choices can influence meta-analysis results, interpretations, and consequent policy decisions. Specifically, use of a more refined approach (e.g., incorporating more use of stratified evaluations and meta-regressions) improves understanding of the data and can help regulators avoid making policy decisions based on erroneous data interpretations. Meta-analyses examining associations between short-term exposures to O_3 and mortality highlight the need to consider publication bias, stratified analyses of seasonal effects, study estimate choices, and multipollutant evaluations. Other examples indicate how advances in analyzing and interpreting data using meta-analysis can result from applying existing tools more effectively as well as developing more sophisticated methodologies. As a general observation, numerous opportunities exist for expanding the use of meta-analysis approaches to more systematically synthesize the diverse, multifaceted scientific literature underlying NAAQS evaluations, even in areas where a number of meta-analyses have been undertaken.

In addition to providing a tool for examining specific studies, meta-analysis can provide a useful framework and impetus for identifying and refining research strategies and for designing more effective and targeted studies. Clearly, use of more consistent and comprehensive research designs and reporting approaches can help mitigate key issues, particularly heterogeneity in study design. Using more consistent and comprehensive study designs can also enhance the usefulness of small studies by providing a way to aggregate such data. Excessive consistency should be avoided as overuse of common designs would complicate evaluation of whether or how study results might differ if certain elements were changed.

As illustrated by the longitudinal studies of children's lead exposures, where researchers coordinated certain aspects of study design and implementation (e.g., Bornschein and Rabinowitz[35]) the goal of improving data synthesis among the studies was hampered because the researchers did not adequately encourage consistency in analyzing the data and reporting results.[68] Similar observations regarding the benefits of greater consistency in study design and data analysis have been made in the more recent scientific literature regarding this research area, e.g., regarding the neurotoxicity literature for lead,[57] and other compounds, such as PCBs.[72] Goodman *et al.*[72] also note that, even as research techniques and test methods evolve (e.g., for exposure and/or outcome assessment, or statistical analyses), studies should continue to include research measures that are comparable to those used in previous studies to provide greater opportunities to compare and synthesize studies conducted at different points in time.

A number of systematic review guidelines for conducting studies or reporting results have also acknowledged the benefits for data synthesis of greater study consistency. One of the seminal efforts to promote more systematic evaluations of human health data, the Cochrane Collaboration, was initiated in

the early 1990s. This international network of individuals and institutions promotes methods and resources for conducting, documenting, and enhancing the accessibility of systematic reviews of randomized control trials of health-care interventions. The *Cochrane Handbook*,[97] which was developed to help scientists conduct credible and comparable clinical trials with humans, provides a consistent approach for conducting clinical studies. Since exposure studies of humans in chambers are similar to clinical studies, adaptation of the principles reflected in this handbook could greatly improve study consistency, which in turn would enhance efforts to compare results. In particular, the handbook encourages researchers to adhere to consistent protocols to reduce the impact of author bias, promote method and process transparency, reduce the potential for duplication, and allow peer review of the planned methods. Another key component is conducting statistical analyses and assessing the quality of the body of evidence. Adoption of such methods would greatly improve analyses and allow more robust interpretations of data generated in chamber studies with air pollutants.

In another effort to promote sound data syntheses, a review by Blair *et al.*[4] discussed uses of meta-analysis techniques for environmental epidemiology data, providing guidelines for when meta-analyses should or should not be used. Two other examples of guideline efforts relevant for evaluations of air pollutants and other contaminants have focused on issues specific for reporting of meta-analysis results. Stroup *et al.*[7] emphasized issues for reporting results from Meta-analysis Of Observational Studies in Epidemiology (MOOSE), while Liberati *et al.*[98] built on a previous effort (the QUality Of Reporting Of Meta-analysis (QUOROM) Statement) to develop guidelines and checklists for reporting systematic reviews and meta-analyses of studies of health-care interventions, as well as in other contexts (the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) Statement). Guidelines have also been developed for reporting laboratory animal meta-analyses,[77] as well as for reporting laboratory animal study data to better support meta-analyses and other systematic data reviews.[81–84] These types of approaches list specific elements to be included in meta-analysis documentation, such as study selection criteria, approaches for assessing study bias, and discussion of sensitivity analyses. Applying these approaches would enhance interpretation, synthesis, and understanding of meta-analysis results.

Reviewing air pollutant research also suggests new areas where meta-analysis techniques could be applied. In particular, studies evaluating potential toxicity mechanisms present new opportunities for synthesizing data using meta-analysis. Because some of these research techniques are evolving (e.g., gene expression studies or imaging studies of structural changes in brain morphology), these research areas are just beginning to be considered in causation evaluations for environmental contaminants. As such, opportunities exist to proactively shape this research area to more effectively support data syntheses using meta-analysis techniques. In other areas of research, such as animal toxicity studies, opportunities exist to better coordinate study methodologies and reporting approaches to enhance data syntheses.

Opportunities also exist for extending lessons learned from studies of air pollutants to other research settings, e.g., to identify specific approaches or data elements that should be included in such studies. For example, identifying potential causes of ADHD has been an active research area in recent years, with studies issued assessing potential roles in ADHD for such substances as NO_2,[99] 2,4,6-trichorophenol,[100] organophosphate pesticides,[101] perfluorinated compounds,[102] phthalate metabolites,[103] and a variety of compounds characterized as endocrine disruptors.[104] As research regarding many of these substances is in an early stage, applying more consistent research and reporting approaches (e.g., based on experience gained from studies of lead) could yield data sets for specific compounds that would be more amenable to synthesis and interpretation. These data could also be useful for assessing the relative magnitude of the associations of these substances with ADHD.

Meta-analysis could also be more extensively applied when studying ecological impacts of environmental contaminants. While meta-analysis applications have increased in this area, they have been most prevalent in such fields as evolutionary ecology, community ecology, and conservation ecology.[105–107] Meta-analysis innovations that have been explored include meta-analyses of meta-analyses (e.g., in the area of plant evolutionary ecology)[108] and within-study meta-analyses as a way to more deeply examine findings from studies that have tested hypotheses using several approaches (e.g., observational and experimental) and measurements (e.g., molecular, behavioral, and physiological).[109]

While meta-analysis is not the only way to systematically review research study data (e.g.,

see Rhomberg *et al.*[9]), the rigor, specificity, and transparency of conducting and documenting this type of evaluation present particular benefits, even for situations where data limitations prevent a meta-analysis from being completed. In particular, the meta-analysis process encompasses a more specific, in-depth consideration of which study elements would need to be combined and how those elements would need to be comparable across studies than does a more qualitative systematic review. In specific evaluations, this detailed perspective helps in selecting studies to include in the systematic analyses, e.g., by helping to identify factors influencing study outcomes, study limitations, and appropriate study groupings (e.g., stratified analyses). Meta-analysis also allows weighting of results from studies of varying quality and strength and yields synthesized, well-documented effect estimates reflecting those considerations. When addressing multifaceted (e.g., multidisciplinary) data sets, meta-analyses can help make such data more manageable and understandable. For example, meta-analysis can provide a systematic approach for assessing the findings for each relevant area and combining them to reach soundly supported and transparent, well-documented conclusions. In particular, meta-analyses in such contexts can help focus attention on key literature aspects warranting additional research or policy-oriented evaluation. Because it emphasizes quantitative syntheses, meta-analysis is particularly well suited for assessing the relative importance of endpoints and for identifying data gaps in existing knowledge. Moreover, because meta-analysis techniques have been applied in diverse settings, efforts to draw upon ideas and methods developed in other contexts can foster cross-disciplinary perspectives.[106,109] In addition, ongoing refinements in the use of certain statistical techniques (e.g., meta-regression, hierarchical models, and hybrid approaches) may provide opportunities for conducting meta-analyses of data that may not, at first, appear amenable to the technique.

The meta-analysis perspective could also help enhance the quality of the published literature, and efforts to design and implement new research efforts. For example, if potential meta-analysis applications were more routinely considered during journal publication reviews, publication bias could be reduced (e.g., if negative or statistically insignificant results were more widely recognized as having value in data syntheses). Moreover, increased pervasiveness of a meta-analysis-oriented mindset—focusing on how the pieces of the research puzzle could fit together—would benefit scientific research and data syntheses regardless of whether or not a quantitative meta-analysis is undertaken.

Overall, as illustrated using examples from the criteria air pollutant literature, meta-analysis is a versatile tool that can help researchers more effectively synthesize existing study data of all types and design data collection efforts in a variety of research contexts. Its adaptability to many data types and its ability to aid understanding of complex data sets is particularly attractive in light of the ever-increasing amount of scientific data being generated and needing interpretation. To date, the use of meta-analysis to support policy determinations has yet to reach its full potential. Opportunities exist for conducting more informative analyses using existing data, for designing studies to better support future data syntheses, and for basing regulatory limits and other science-based policy decisions on more representative analyses. In particular, thoughtful use of meta-analysis shows much promise to support determinations that must integrate information from many disciplines. The case studies drawn from scientific and regulatory evaluations of criteria air pollutants yield observations broadly applicable to a wide range of research and policy areas. In particular, the observations presented in this article can help to inform use of meta-analysis within focused research contexts or where one or more meta-analyses can be combined to support evaluations of a more multifaceted issue.

ACKNOWLEDGMENTS

This work was funded by the Texas Commission on Environmental Quality (TCEQ). This commentary represents the individual professional views of the authors and not necessarily the views of TCEQ.

REFERENCES

1. Egger M, Smith GD, Altman D. Systematic Reviews in Health Care: Meta-Analysis in Context, 2nd ed. London: BMJ Publishing Group, 2001.
2. Cochrane WG. The comparison of percentages in matched samples. Biometrika, 1950; 37(3/4):256–266.
3. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine 2002; 21(11):1539–1558.
4. Blair A, Burg J, Foran J, Gibb H, Greenland S, Morris R, Raabe G, Savitz D, Teta J, Wartenberg D, Wong O, Zimmerman R. Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. Regulatory Toxicology and Pharmacology 1995; 22(2):189–197.

5. Nordmann AJ, Kasenda B, Briel M. Meta-analyses: What they can and cannot do. Swiss Medical Weekly 2012; 142:w13518.
6. Greenland S, O'Rourke K. Meta analysis. Pp. 652–683 in Rothman KJ, Greenland S, Lash TL, (eds). Modern Epidemiology, 3rd ed. Philadelphia: Lippincott, Williams & Wilkins; 2008.
7. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. JAMA. 2000; 283:2008–2012.
8. Centre for Reviews and Dissemination (CRD). Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care, 2009 January. Available at: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
9. Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, Honeycutt M, Kaminski NE, Paoli G, Pottenger LH, Scherer RW, Wise KC, Becker RA. A survey of frameworks for best practices in weight-of-evidence analyses. Critical Reviews in Toxicology 2013; 43(9):753–784.
10. US EPA. Clean Air Act, 2012 Feb 17. Available at: http://www.epa.gov/air/caa/.
11. US EPA. Integrated Science Assessment for Carbon Monoxide. National Center for Environmental Assessment (NCEA) – RTP Division, EPA/600/R-09/019F. 2010 Jan.
12. US EPA. Integrated Science Assessment for Lead (Final). National Center for Environmental Assessment (NCEA), EPA/600/R-10/075F. 2013 June.
13. US EPA. Integrated Science Assessment for Ozone and Related Photochemical Oxidants (Final). National Center for Environmental Assessment (NCEA), EPA/600/R-10/076F. 2013 Feb.
14. US EPA. 2009a. Integrated Science Assessment for Particulate Matter (Final). Office of Research and Development, National Center for Environmental Assessment (NCEA) – RTP Division, EPA/600/R-08/139F. 2009 Dec.
15. US EPA. Nitrogen Dioxide (NO_2) Primary Standards, 2013 July 12. Available at: http://www.epa.gov/ttn/naaqs/standards/nox/s_nox_index.html.
16. US EPA, Integrated Science Assessment for Oxides of Nitrogen—Health Criteria (Second External Review Draft). National Center for Environmental Assessment (NCEA), EPA/600/R-14/006. 2015 Jan. Available at: http://www.epa.gov/ncea/isa.
17. US EPA. Sulfur Dioxide (SO_2) Primary National Ambient Air Quality Standards. 2013 July 12. Available at: http://www.epa.gov/ttn/naaqs/standards/so2/s_so2_index.html.
18. US EPA. Integrated Science Assessment for Oxides of Nitrogen. National Center for Environmental Assessment (NCEA) – RTP Division, EPA/600/R-08/071. 2008 July.
19. Goodman JE, Chandalia JK, Thakali S, Seeley M. Meta-analysis of nitrogen dioxide exposure and airway hyperresponsiveness in asthmatics. Critical Reviews in Toxicology 2009; 39(9):719–742.
20. Hill AB. The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine, 1965; 58(5):295–300.
21. McDonnell WF. Utility of controlled human exposure studies for assessing the health effects of complex mixtures and indoor air pollutants. Environmental Health Perspectives, 1993; 101(Suppl. 4):199–203.
22. US EPA. Primary national ambient air quality standards for nitrogen dioxide (final rule). 2010 Feb 9; Federal Register, 2010; 75(26):6473–6538.
23. Cockcroft DW, Davis BE, Boulet LP, Deschesnes F, Gauvreau GM, O'Byrne PM, Watson RM. The links between allergen skin test sensitivity, airway responsiveness and airway response to allergen. Allergy, 2005; 60(1):56–59.
24. Cockcroft DW, Davis BE. Mechanisms of airway hyperresponsiveness. Journal of Allergy and Clinical Immunology, 2006; 118(3):551–559.
25. US EPA. Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information (OAQPS Staff Paper). Office of Air Quality Planning and Standards, EPA-452/R-07-007. 2007 July.
26. US EPA. Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard: Final Report. Office of Air Quality Planning and Standards, EPA-452/R-09-007. 2009 July.
27. Goodman JE, Dodge DG, Bailey LA. A framework for assessing causality and adverse effects in humans with a case study of sulfur dioxide. Regulatory Toxicology and Pharmacology, 2010; 58:308–28.
28. Adams WC. Comparison of chamber and face-mask 6.6-h exposures to ozone on pulmonary function and symptoms responses. Inhalation Toxicology, 2002; 14(7):745–764.
29. Adams WC. Comparison of chamber 6.6-h exposures to 0.04–0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. Inhalation Toxicology, 2006; 18(2):127–136.
30. Schelegle ES, Morales CA, Walby WF, Marion S, Allen RP. 6.6-h inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. American Journal of Respiratory and Critical Care Medicine, 2009; 180(3):265–272.
31. Kim CS, Alexis NE, Rappold AG, Kehrl H, Hazucha MJ, Lay JC, Schmitt MT, Case M, Devlin RB, Peden DB, Diaz-Sanchez D. Lung function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 h. American Journal of Respiratory and Critical Care Medicine, 2011; 183:1215–1221.
32. Dominici F, Sheppard L, Clyde M. Health effects of air pollution: A statistical review. International Statistical Review, 2003; 71:243–276.
33. Jerrett M, Burnett RT, Pope CA, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M. Long-term ozone exposure and mortality. New England Journal of Medicine, 2009; 360(11):1085–1095.
34. Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope A, Thurston G, Calle EE, Thun MJ. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Research Report 140. Boston, MA: Health Effects Institute (US), 2009 May.
35. Bornschein RL, Rabinowitz MB, Eds. The Second International Conference on Prospective Studies of Lead, Cincinnati, Ohio, April 9–11, 1984. Environmental Research, 1985; 38(1).
36. Stieb DM, Judek S, Burnett RT. Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season. Journal of the Air & Waste Management Association, 2002; 52:470–484.
37. Ostro B. Outdoor air pollution: Assessing the environmental burden of disease at national and local levels. World Health Organization Environmental Burden of Disease Series No. 5. Geneva, World Health Organization, 2004 Available at: http://www.who.int/quantifying_ehimpacts/publications/ebd5.pdf?ua=1
38. Bell ML, Dominici F, Samet JM. A meta-analysis of time-series studies of ozone and mortality with comparison to the National Morbidity, Mortality, and Air Pollution Study. Epidemiology, 2005; 16(4):436–445.
39. Anderson GB, Bell ML. Does one size fit all? The suitability of standard ozone exposure metric conversion ratios and implications for epidemiology. Journal of Exposure Science and Environmental Epidemiology, 2010; 20(1):2–11.

40. Samet JM, Zeger SI, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity, mortality, and air pollution in the United States. HEI Research Report 94, Part II. Cambridge, MA: Health Effects Institute (US), 2000 June.
41. Bell ML, Kim JY, Dominici F. Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies. Environmental Health Perspectives, 2007; 115(11):1591–1595.
42. Bell ML, Dominici F. Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities. American Journal of Epidemiology, 2008; 167(8):986–997.
43. Smith RL, Xu B, Switzer P. Reassessing the relationship between ozone and short-term mortality in U.S. urban communities. Inhalation Toxicology, 2009; 21(Suppl. 2):37–61.
44. Stylianou M, Nicolich MJ. Cumulative effects and threshold levels in air pollution mortality: Data analysis of nine large US cities using the NMMAPS dataset. Environmental Pollution, 2009; 157(8–9):2216–2223.
45. Schwartz J. Air pollution and daily mortality: A review and meta analysis. Environmental Research, 1994; 64:36–52.
46. Stieb DM, Judek S, Burnett RT. Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models. Journal of the Air & Waste Management Association, 2003; 53(3):258–261.
47. Ito K, De Leon SF, Lippmann M. Associations between ozone and daily mortality: Analysis and meta-analysis. Epidemiology, 2005; 16(4):446–457.
48. Levy JI, Chemerynski SM, Sarnat JA. Ozone exposure and mortality: An empiric Bayes metaregression analysis. Epidemiology, 2005; 16(4):458–468.
49. Kaufman AS. Do low levels of lead produce IQ loss in children? A careful examination of the literature. Archives of Clinical Neuropsychology, 2001; 16(4):303–341.
50. Goodlad JK, Marcus DK, Fulton JJ. Lead and attention-deficit/hyperactivity disorder (ADHD) symptoms: A meta-analysis. Clinical Psychology, 2013; 33(3):417–425
51. Navas-Acien A, Schwartz BS, Rothenberg SJ, Hu H, Silbergeld EK, Guallar E. Bone lead levels and blood pressure endpoints: A meta-analysis. Epidemiology, 2008; 19(3):496–504.
52. Fu H, Boffetta P. Cancer and occupational exposure to inorganic lead compounds: a meta-analysis of published data. Occupational and Environmental Medicine, 1995; 52(2):73–81.
53. Scinicariello F, Murray HE, Moffett DB, Abadin HG, Sexton MJ, Fowler BA. Lead and delta-aminolevulinic acid dehydratase polymorphism: Where does it lead? A meta-analysis. Environmental Health Perspectives, 2007; 115(1): 35–41.
54. Yeoh B, Woolfenden S, Lanphear B, Ridley GF, Livingstone N. Household interventions for preventing domestic lead exposure in children (review). Cochrane Database, 2012. Syst. Rev. 4:CD006047.
55. Marcus DK, Fulton JJ, Clarke EJ. Lead and conduct problems: A meta-analysis. Journal of Clinical Child & Adolescent Psychology, 2010; 39(2):234–241.
56. Anderson HR, Favarato G, Atkinson RW. Long-term exposure to air pollution and the incidence of asthma: Meta-analysis of cohort studies. Air Quality, Atmosphere, and Health, 2013; 6(1):47–56.
57. Bellinger DC. Interpreting epidemiologic studies of developmental neurotoxicity: Conceptual and analytic issues. Neurotoxicology and Teratology, 2009; 31(5):267–274.
58. Stewart P. The critical concept of control in human neurobehavioral toxicology studies: We can, and must, do better. Pp. 361–378 in Bellinger DC, (ed). Human Developmental Neurotoxicology. New York: Taylor & Francis Group, 2006. p. 361–378.
59. Meyer-Baron M, Seeber A. A meta-analysis for neurobehavioral results due to occupational lead exposure with blood lead concentrations <70 microg/100 ml. Archives of Toxicology, 2000; 73(10/11):510–518.
60. Goodman M, LaVerda N, Mandel J. Commentary on "A meta-analysis for neurobehavioral results due to occupational lead exposure with blood lead concentrations <70 micrograms/100 ml." Archives of Toxicology, 2001; 75(7):439–440.
61. Goodman M, LaVerda N, Clarke C, Foster ED, Iannuzzi J, Mandel J. Neurobehavioral testing in workers occupationally exposed to lead: Systematic review and meta-analysis of publications. Occupational and Environmental Medicine, 2002; 59(4):217–223.
62. Schwartz BS, Stewart W, Hu H. Neurobehavioral testing in workers occupationally exposed to lead [letter]. Occupational and Environmental Medicine, 2002; 59:648–650.
63. Seeber A, Meyer-Baron M, Schaper M. A summary of two meta-analyses on neurobehavioral effects due to occupational lead exposure. Archives of Toxicology, 2002; 76(3):137–145.
64. Seeber A, Meyer-Baron M. Neurobehavioral testing in workers occupationally exposed to lead [letter and authors' reply]. Occupational and Environmental Medicine, 2003; 60:145–146.
65. Clean Air Science Advisory Committee (CASAC). Letter to: B Perciasepe (US EPA). 2013 Jun 4. Located at: Science Advisory Board, EPA-CASAC-13–004. Re: CASAC Review of EPA's "Integrated Science Assessment for Lead (Third External Review Draft – November 2012)."
66. Mazumdar M, Bellinger DC, Gregas M, Abanilla K, Bacic J, Needleman HL. Low-level environmental lead exposure in childhood and adult intellectual function: A follow-up study. Environmental Health, 2011; 10:24.
67. Braun JM, Hoffman E, Schwarz J, Sanchez B, Schnaas L, Mercado-Garcia A, Solano-Gonzalez M, Bellinger DC, Lanphear BP, Hu H, Tellez-Rojo MM, Wright RO, Hernandez-Avila M. Assessing windows of susceptibility to lead-induced cognitive deficits in Mexican children. Neurotoxicology, 2012; 33(5):1040–1047.
68. Thacker SB, Hoffman DA, Smith J, Steinberg K, Zack M. Effect of low-level body burdens of lead on the mental development of children: Limitations of meta-analysis in a review of longitudinal data. Archives of Environmental Health, 1992; 47(5):336–346.
69. Lanphear BP, Hornung R, Khoury J, Yolton K, Baghurst P, Bellinger DC, Canfield RL, Dietrich KN, Bornschein R, Greene T, Rothenberg SJ, Needleman HL, Schnaas, L, Wasserman G, Graziano J, Roberts R. Low-level environmental lead exposure and children's intellectual function: An international pooled analysis. Environmental Health Perspectives, 2005; 113(7):894–899.
70. Bellinger DC. Effect modification in epidemiologic studies of low-level neurotoxicant exposures and health outcomes. Neurotoxicology and Teratology, 2000; 22(1):133–140.
71. Bellinger DC. Lead neurotoxicity in children: Decomposing the variability in dose-effect relationships. American Journal of Industrial Medicine 2007; 50(10):720–728.
72. Goodman M, Squibb K, Youngstrom E, Anthony LG, Kenworthy L, Lipkin PH, Mattison DR, Lakind JS. Using systematic reviews and meta-analyses to support regulatory decision-making for neurotoxicants: Lessons learned from a case study of PCBs. Environmental Health Perspectives, 2010; 118:727–734.
73. Meyer-Baron M, Schaper M, Knapp G, Lucchini R, Albini E, Bast-Pettersen R, He S, Yuan H, Niu Q, Wang XL, Yang YJ, Iregren A, Sjogren B, Blond M, Laursen P, Netterstrom

B, Mergler D, Bowler R, van Thriel C. Statistical means to enhance the comparability of data within a pooled analysis of individual data in neurobehavioral toxicology. Toxicology Letters, 2011; 206(2):144–151.
74. Levy JI, Fabian MP, Peters JL. Meta-analytic approaches for multi-stressor dose-response function development: Strengths, limitations, and case studies. 2013 August.
75. Shin HH, Cohen A, Pope CA III, Ezzati M, Lim SS, Hubbell B, Burnett RT. Critical issues in combining disparate sources of information to estimate the global burden of disease attributable to ambient fine particulate matter exposure, Methods for Research Synthesis: A Cross-Disciplinary Workshop, 2013 August.
76. Glass GV. Primary, secondary, and meta-analysis of research. Educational Research, 1976; 5(10):3–8.
77. Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR. A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting. Journal of Environmental Science and Health, B., 2006; 41(7):1245–1258.
78. Valberg PA, Crouch EAC. Meta-analysis of rat lung tumors from lifetime inhalation of diesel exhaust. Environmental Health Perspectives, 1999; 107(9):693–699.
79. Chang X, Zhang Y, Tang M, Wang B. Health effects of exposure to nano-TiO2: A meta-analysis of experimental studies. Nanoscale Research Letters, 2013; 8(1):51.
80. Tseng GC, Ghosh D, Feingold E. Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Research, 2012; 40(9):3785–3799.
81. Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, Buchan A, van der Worp HB, Traystman RJ, Minematsu K, Donnan GA, Howells DW. Reprint: Good laboratory practice: Preventing introduction of bias at the bench. Journal of Cerebral Blood Flow & Metabolism, 2009; 29(2):221–223.
82. Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three Rs, and to make systematic reviews more feasible. Alternatives to Laboratory Animals, 2010; 38(2):167–182.
83. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biology, 2010; 8(6):e1000412.
84. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, Macleod MR. Can animal models of disease reliably inform human studies? PLoS Medcine, 2010; 7(3):e1000245.
85. Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: A systematic review. Environmental Health Perspectives, 2013; 121(9):985–992.
86. Pennings JL, Kimman TG, Janssen R. Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques. PLoS ONE, 2008; 3(7):e2596.
87. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Medicine, 2008; 5(9):e184.
88. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)—Toward standards for microarray data. Nature Genetics, 2001; 29(4):365–371.
89. Schmidberger M, Lennert S, Mansmann U. Conceptual aspects of large meta-analyses with publicly available microarray data: A case study in oncology. Bioinformatics and Biology Insights, 2011; 5:13–39.
90. Kristiansson E, Osterlund T, Gunnarsson L, Arne G, Joakim Larsson DG, Nerman O. A novel method for cross-species gene expression analysis. BMC Bioinformatics, 2013; 14:70.
91. Nakao T, Radua J, Rubia K, Mataix-Cols D. Gray matter volume abnormalities in ADHD: Voxel-based meta-analysis exploring the effects of age and stimulant medication. American Journal of Psychiatry, 2011; 168(11):1154–1163.
92. Brubaker CJ, Schmithorst VJ, Haynes EN, Dietrich KN, Egelhoff JC, Lindquist DM, Lanphear BP, Cecil KM. Altered myelination and axonal integrity in adults with childhood lead exposure: A diffusion tensor imaging study. Neurotoxicology, 2009; 30(6):867–875.
93. Brubaker CJ, Dietrich KN, Lanphear BP, Cecil KM. The influence of age of lead exposure on adult gray matter volume. Neurotoxicology, 2010; 31(3):259–266.
94. Cecil KM, Brubaker CJ, Adler CM, Dietrich KN, Altaye M, Egelhoff JC, Wessel S, Elangovan I, Hornung R, Jarvis K, Lanphear BP. Decreased brain volume in adults with childhood lead exposure. PLoS Medicine, 2008; 5(5):e112.
95. Cecil KM, Dietrich KN, Altaye M, Egelhoff JC, Lindquist DM, Brubaker CJ, Lanphear BP. Proton magnetic resonance spectroscopy in adults with childhood lead exposure. Environmental Health Perspectives, 2011; 119(3):403–408.
96. Texas Commission on Environmental Quality (TCEQ). Nickel and Inorganic Nickel Compounds (Final). 2011 June 1. Available at: http://www.tceq.state.tx.us/assets/public/implementation/tox/dsd/final/june11/nickel_&_compounds.pdf.
97. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0). The Cochrane Collaboration, 2011 March. Available at: http://www.cochrane-handbook.org/.
98. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gatzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care intverventions: Explanation and elaboration. Annals of Internal Medicine, 2009; 151:W65–W94.
99. Morales E, Julvez J, Torrent M, de Cid R, Guxens M, Bustamante M, Kunzli N, Sunyer J. Association of early-life exposure to household gas appliances and indoor nitrogen dioxide with cognition and attention behavior in preschoolers. American Journal of Epidemiology, 2009; 169(11):1327–1336.
100. Xu X, Nembhard WN, Kan H, Kearney G, Zhang ZJ, Talbott EO. Urinary trichlorophenol levels and increased risk of attention deficit hyperactivity disorder among US school-aged children. Occupational and Environmental Medicine, 2011; 68(8):557–561.
101. Bouchard MF, Bellinger DC, Wright RO, Weisskopf MG. Attention-deficit/hyperactivity disorder and urinary metabolites of organophosphate pesticides. Pediatrics, 2010; 125(6):e1270–e1277.
102. Gump BB, Wu Q, Dumas AK, Kannan K. Perfluorochemical (PFC) exposure in children: Associations with impaired response inhibition. Environmental Science & Technology, 2011; 45(19):8151–8159.
103. Engel SM, Miodovnik A, Canfield RL, Zhu C, Silva MJ, Calafat AM, Wolff MS. Prenatal phthalate exposure is associated with childhood behavior and executive functioning. Environmental Health Perspectives, 2010; 118(4):565–571.
104. de Cock M, Maas YG, van de Bor M. Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review. Acta Paediatrica, 2012; 101(8):811–818.
105. Gurevitch J, Curtis PS, Jones MH. Meta-analysis in ecology. Advances in Ecological Research, 2001; 32:199–247.

106. Gates S. Review of methodology of quantitative reviews using meta-analysis in ecology. Journal of Animal Ecology, 2002; 71(4):547–557.
107. Nakagawa S, Poulin R. Meta-analytic insights into evolutionary ecology: An introduction and synthesis. Evolutionary Ecology, 2012; 26(5):1085–1099.
108. Castellanos MC, Verdu M. Meta-analysis of meta-analyses in plant evolutionary ecology. Evolutionary Ecology, 2012; 26(5):1187–1196.
109. Nakagawa S, Santos ESA. Methodological issues and advances in biological meta-analysis. Evolutionary Ecology, 2012; 26(5):1253–1274.