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

Julie E. Goodman*, Catherine Petito Boyce, Sonja N. Sax, Leslie A. Beyer, and Robyn L. Prueitt
(Gradient, Cambridge, MA)

Working Paper prepared for:

Methods for Research Synthesis: A Cross-Disciplinary Workshop

Harvard Center for Risk Analysis

October 3, 2013

www.hcra.harvard.edu

***Corresponding author: jgoodman@gradientcorp.com**

Acknowledgements: This paper was funded by the Texas Commission on Environmental Quality (TCEQ). The authors have sole responsibility for the writing and contents of this paper. This paper represents the professional opinions of the authors and not necessarily those of TCEQ.

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

Julie E. Goodman, Catherine Petito Boyce, Sonja N. Sax, Leslie A. Beyer, and Robyn L. Prueitt
Gradient, Cambridge, MA

Abstract

Meta-analyses offer a rigorous and transparent systematic framework for data synthesis 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 insights to support scientific understanding and policy decision-making. A variety of meta-analysis applications can be illustrated by evaluations that have been or could be conducted in the context of the National Ambient Air Quality Standards for criteria air pollutants, for which US EPA and others have conducted meta-analyses on data from epidemiology and controlled exposure studies. These evaluations demonstrate the strengths and limitations of meta-analysis, issues that arise in addressing different data categories, how the choices made in conducting a meta-analysis can influence the interpretation of results, and how meta-analysis can be used to address bias and heterogeneity. Meta-analyses have not been used as extensively to evaluate toxicity and mechanistic studies, but such analyses could be beneficial. 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 be beneficial to scientific research and data syntheses regardless of whether the data are applied in a quantitative meta-analysis or not. 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 larger question where different kinds of data need to be brought to bear. This is not only true for air pollution data, but for any scientific or policy area where information from a variety of disciplines must be considered to address a broader research question.

Key words: Meta-analysis, air pollutants, data synthesis, bias, heterogeneity.

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 datasets addressing a specific research question to yield an overall "consensus" of the data (Egger *et al*., 2001). A well-conducted meta-analysis prepared following the steps illustrated in Figure 1.1 incorporates a number of key features that can help minimize bias, random errors, and subjectivity in data evaluations (CRD, 2009). These features include requirements for 1) a thorough literature search; 2) clear and transparent eligibility criteria for selecting studies to include in the analyses; 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 heterogeneity and bias. While some of these features can be incorporated into more qualitative, narrative systematic study reviews (reviewed by Rhomberg *et al*., 2013), the more rigorous, quantitative perspective on the study data 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 is limited to evaluating only 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 require a comprehensive analysis incorporating data from multiple disciplines.

Step 1: Identify the need for a review

Step 2: Develop review protocol

Step 3: Develop literature search strategy

Step 4: Study selection

Step 5: Data extraction

Step 6: Quality assessment

Step 7: Data synthesis

Step 8: Report and recommendations

Step 9: Apply evidence to practice

Figure 1.1 Steps of a Meta-analysis. Adapted from CRD, 2009.

Although data from individual studies can be combined to form one dataset and analyzed in what is called a pooled analysis, 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 (Blair *et al*., 1995). In addition, combining results of individual studies can make them more generally applicable (*e.g.*, across various populations) (Nordmann *et al*., 2012). 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).

A variety of meta-analysis applications can be illustrated by evaluations that have been or could be conducted in the context of developing the National Ambient Air Quality Standards (NAAQS) as mandated by the Clean Air Act (US EPA, 2012). Reviewing data from a variety of disciplines (including epidemiology, toxicology, atmospheric science, and exposure science), US EPA develops 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-of-evidence (WoE) for causal determinations for each air pollutant and various human health and ecological outcomes. For each substance, US EPA documents these evaluations in an Integrated Science Assessment (ISA) report. Final ISAs have been completed for four of the criteria pollutants: CO (US EPA, 2010a), lead (US EPA, 2013a), O_3 (US EPA, 2013b), and PM (US EPA, 2009a). The current review processes for NO_2 (US EPA, 2013c) and SO_2 (US EPA, 2013d) 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, 2008; Goodman *et al*., 2009). 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 paper 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 meta-analyses that have been conducted and identify opportunities for refinements to existing meta-analyses or expanding 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 the results of individual meta-analyses, and 3) how meta-analysis considerations can enhance the design and implementation of new research efforts. We also discuss how the results of separate meta-analyses can be brought together to address a larger question (*e.g.*, causation determination) where findings from a variety of data types and research areas need to be integrated.

2 Controlled Exposure Studies

In controlled exposure studies, people with regulated activity levels are exposed to known concentrations of substances, such as air pollutants, under carefully controlled environmental conditions in exposure chambers (US EPA, 2008). This exposure method minimizes possible confounding by other factors, and sensitive experimental techniques can be used to measure health effects (and markers of injury) 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 that are easily measured and can be described as categorical or continuous variables (McDonnell, 1993). 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 (particularly those of criteria pollutants) have a number of features that affect the interpretation of results (US EPA, 2008). 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 the health effects of exposure 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 effects seen 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 question. In fact, the small number of subjects per study makes meta-analysis a very important tool to evaluate these studies as a whole.

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 (US EPA, 2008). US EPA classified individuals as either having 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 second (FEV_1)] and 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 (US EPA, 2008).

Altogether, the analysis included data from 17 studies of 355 asthmatics with 1-hour exposures ranging from 0.1-0.6 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-minute exposures to $\geq$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 (2010b) stated that it was appropriate to consider NO_2-induced AHR in characterizing NO_2-associated health risks; it based the 1-hour NAAQS of 0.1 ppm largely on this analysis, its evaluation of pertinent observational epidemiology literature, and consideration of a potential shift in the distribution of health effects of NO_2 at the population level.

US EPA reported heterogeneity in the responses among 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.* non-obstructed asthmatics, and varying use of medications by participants) (US EPA, 2008). 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 the quantitative relationship between NO_2 exposures and several specific AHR measures and particular factors influencing exposure-response associations (Goodman *et al.*, 2009). Similar to US EPA, Goodman *et al*. (2009) evaluated the effects of NO_2 exposure (at concentrations ranging from 0.1-0.6 ppm) on AHR to airway challenges in a total of 570 asthmatics in 28 controlled exposure studies (Goodman *et al.*, 2009). This meta-analysis included studies of both specific and nonspecific responsiveness, and it stratified analyses by airway challenge (specific/nonspecific), exposure method (mouthpiece/whole chamber), and activity during exposure (rest/exercise), as all of these factors have been demonstrated to affect AHR (Cockcroft *et al*., 2005; Cockcroft and Davis, 2006).

The primary difference between the US EPA (2008) and Goodman *et al*. (2009) analyses is that while US EPA (2008) evaluated only the percent of people with decreased AHR, Goodman *et al*. (2009) also evaluated the magnitude of the change in AHR following NO_2 (*vs.* filtered air) exposure. Magnitude was assessed 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.*, Figure 2.1). Goodman *et al*. (2009) 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 (2007, 2009b) criteria for what constitutes moderate or severe outcomes. Exposure-response associations – assessed by meta-regression and evaluating effects stratified by exposure level (0.1-0.2, 0.2-0.3 ppm, *etc.*) for the fraction of asthmatics with greater AHR following NO_2 exposure – were not statistically significant in the overall or stratified analyses (*e.g.*, Figure 2.2).

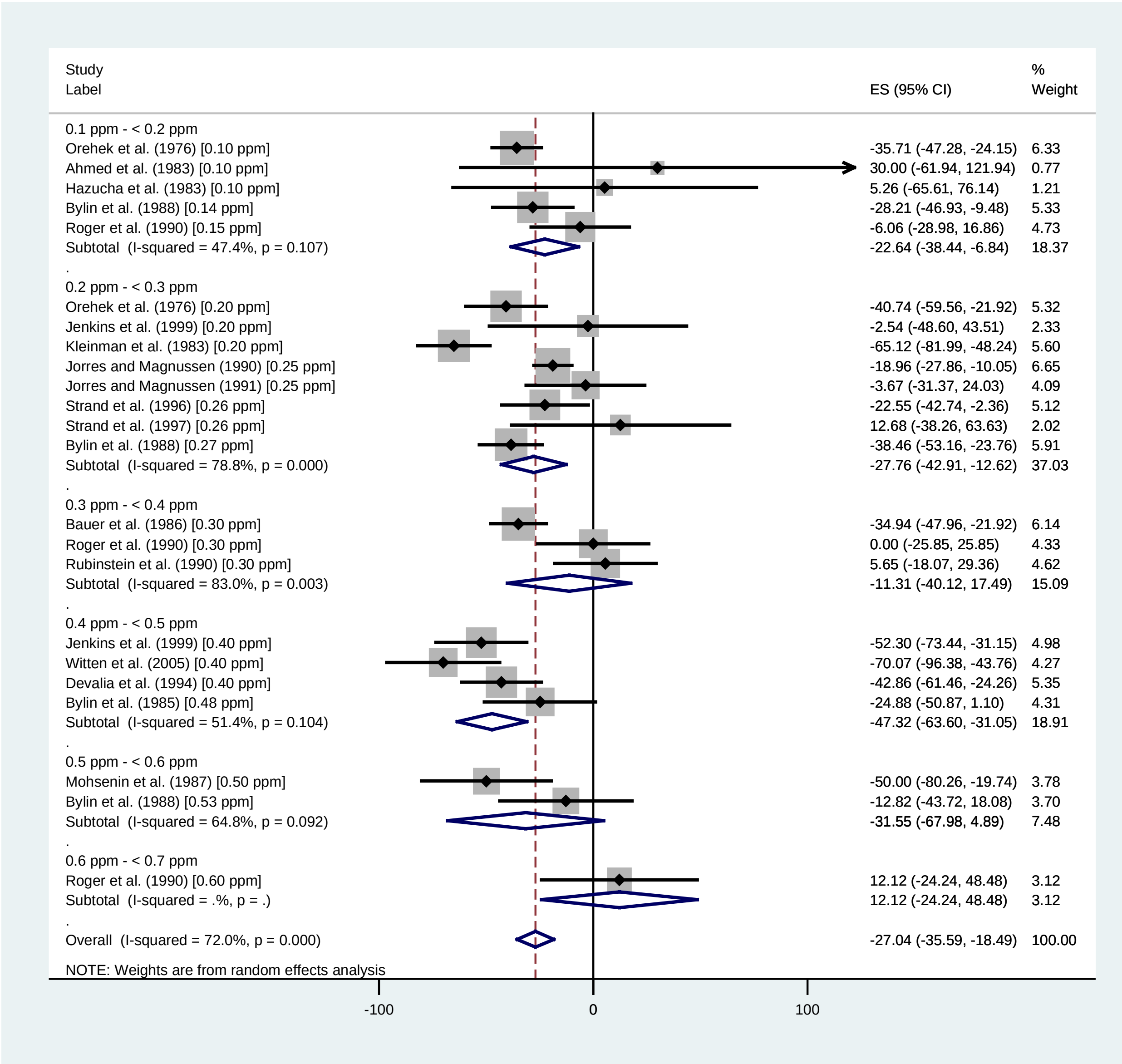

Figure 2.1 Forest Plot Showing the Difference in Responses to Airway Challenge Provocative Doses Following Exposure of Asthmatics to NO_2 *vs.* Air. This figure illustrates the types of meta-analysis findings that can be graphically illustrated in a forest plot, *e.g.*, the average change per dose from 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 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 such observations as whether results are consistent across studies or whether a dose-response relationship appears to exist. Adapted from Goodman *et al*. (2009).

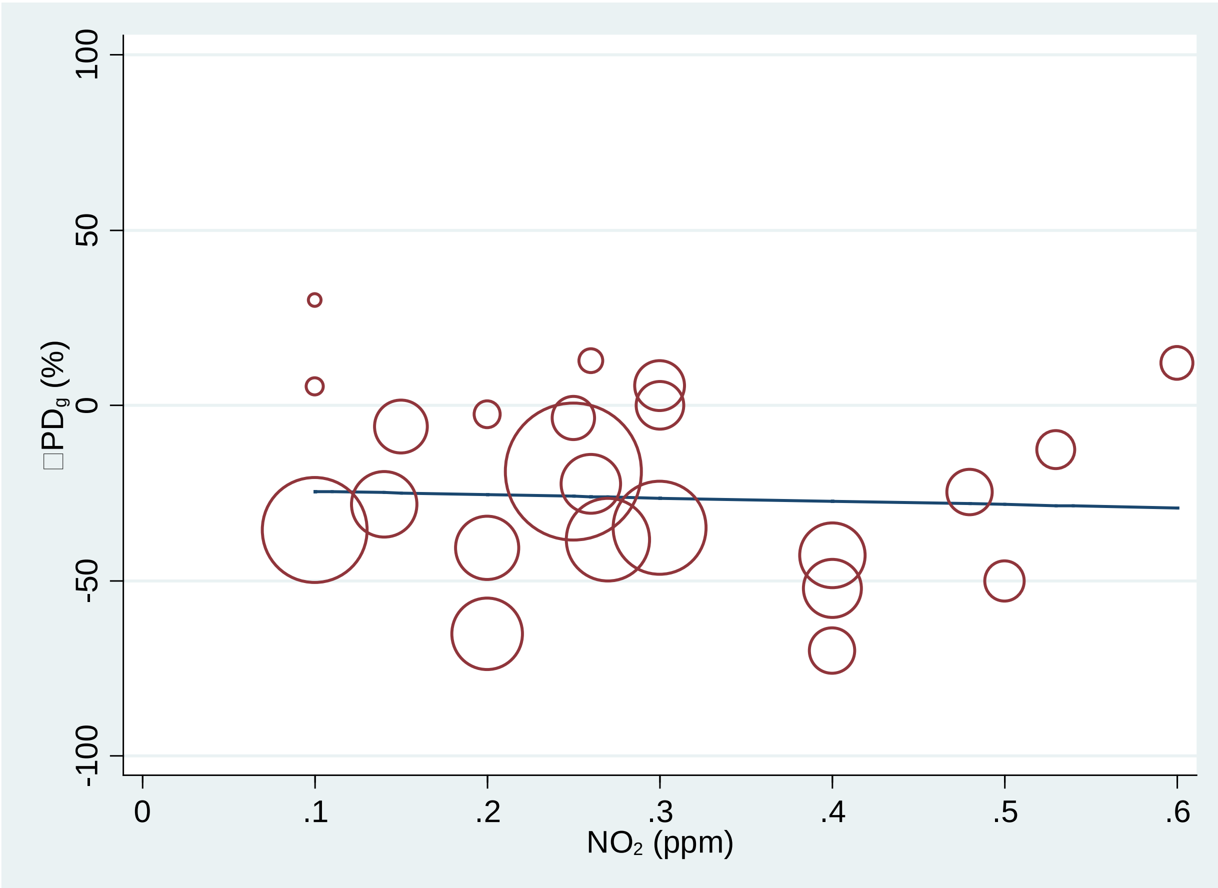

Figure 2.2 Association Between NO_2 Exposure and Airway Hyper-responsiveness in Asthmatics Based on Meta-regressions for the Difference Between Airway Challenge Provocative Dose Following Exposure to NO_2 *vs.* Air. This figure illustrates the use of a bubble plot to display meta-regression results. 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*. (2009).

Compared to the Goodman *et al*. (2009) methodology, the US EPA (2008) approach was able to combine 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, such as FEV_1) 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, the US EPA (2008) approach did not provide a means to evaluate the magnitude of effect or whether the magnitude of effect increased as exposure increased (*i.e.*, exposure-response). As a result, US EPA (2008) was unable to fully evaluate causation or assess whether effects were indicative of homeostasis or more likely to indicate adversity. As 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 have the potential to play an important role in quantifying protective standards. In particular, meta-regression is a powerful tool that has been underutilized to date; as illustrated above, greater use of meta-regression and more refined meta-analysis approaches in interpreting data from controlled exposure studies would provide greater insights into the quantitative nature of exposure-response relationships, as well as the key factors influencing such relationships. Such insights are particularly helpful when assessing whether observed statistically significant associations are causal and whether the observed effects constitute adverse impacts.

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 (*e.g.*, Adams, 2002, 2006; Schelegle *et al*., 2009; Kim *et al*., 2011), the majority of which reported no statistically significant changes after 6.6-hour exposures (with moderate exercise) to up to 0.06 ppm O_3. Because they are fairly homogeneous, in theory, a meta-analysis could be used to increase the power to detect whether there are statistically significant effects. Unfortunately, the publications from these studies do not provide sufficient data to do so.

In these studies, healthy young adults were exposed to O_3 while exercising for up to 6.6 hours. Their lung functions were measured at several time points; however, not all publications of these studies provided effects data for each time point. Omitting the data from the intermediate time points complicates interpretation of the results (*e.g.*, by preventing researchers from evaluating the possibility of false positive findings). There is always a chance that a statistical comparison will make it appear that a true difference exists when, in fact, it does not (*i.e.*, a false positive), and the chance of obtaining a false positive result increases with an increased number of statistical comparisons. Similarly, in the absence of complete data, scientists conducting meta-analyses cannot determine 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 the 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, the reporting choices in the original study and the meta-analysis design can affect the interpretation of findings. For example, if the authors of individual studies provide only a subset of the available data, subsequent syntheses of those data will be hampered and may yield biased results. Moreover, even when more complete information is available, meta-analyses of the same or similar studies on a specific topic can be interpreted differently, depending on how the meta-analyses are conducted. Important factors that can affect interpretation include which studies are included in the analysis, the overall size of the dataset, the specific exposure conditions that are evaluated, how outcomes are measured, how an outcome is considered in the analysis (*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, the use of 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 amongst studies in the data that are collected, analyzed, and reported. Such differences can affect choices that must be 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 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). Efforts to synthesize findings across observational studies are 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 makes it challenging to synthesize 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 study results (*e.g.*, Stroup *et al*., 2000). In addition, when data from observational studies are applied in decision-making contexts, their heterogeneous nature presents more choices in selecting specific research areas warranting synthesis, more need to determine areas where information synthesis will have the greatest impact on policy decision-making, and more potential benefit in making the available literature more understandable.

One aspect of the heterogeneity of observational epidemiology studies is the range of study designs that have been applied, including time-series, cross-sectional, cohort, case-control, case-crossover, and panel studies. These study designs have been used 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 data from fixed-site air monitors located across the US used to assess regulatory compliance. Study designs can also vary with respect to exposure measures, *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. In these studies, daily population-averaged pollution exposure estimates are compared with daily population-averaged health outcome event 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 (GAM) or Generalized Linear Models (GLMs), two statistical modeling approaches that differ in their degree of flexibility and how confounding factors (such as effects of seasonality trends and weather variables) can be addressed (Dominici *et al.*, 2003). Different effect estimates have been observed 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 (*e.g.*, Jerrett *et al*., 2009;Krewski *et al*., 2009). Inferences from these studies are based on differences in pollution levels between cities, as opposed to day-to-day differences in pollution levels in a single city. Any factor that varies from city to city could be a potential confounder, including socioeconomic and lifestyle factors, making controlling for confounding 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). Cohort studies of lead effects have also been conducted, including a set of studies initiated in the early 1980s that were conducted with some degree of coordination to enhance comparisons and synthesis of results (US EPA, 2013a; Bornschein and Rabinowitz, 1985).

For a number of criteria air pollutants (*e.g.*, PM, SO_2 and NO_2), US EPA has focused its evaluations on health effects related to 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 (US EPA, 2009a). 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 (US EPA, 2013a).

An important source of heterogeneity is in approaches used to estimate exposure. 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. Some researchers have applied conversion factors (*e.g.*, to convert PM_{10} to $PM_{2.5}$) to address this problem (*e.g.*, Stieb *et al*. 2002).

Heterogeneity can also arise from the way in which exposure duration is averaged; *e.g.*, health effects of O_3 exposures have been evaluated based on a daily average, a daily maximum 8-hr average, or a daily 1-hr 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 (*e.g.*, Bell *et al*., 2005). These conversion factors, however, have been shown to introduce error and distort observed pollution patterns, and they can result in biased health effect estimates (Anderson and Bell, 2010).

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 multi-step and multi-faceted 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 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 multi-city studies, where a common analytic framework is used for estimating city-specific effects. Applying similar techniques to those used in meta-analysis, these effects are then combined to obtain a summary estimate of effects across all cities.[1] Both approaches are particularly useful in exploring sources and impacts of heterogeneity, increasing statistical power to detect effects, yielding overall effect estimates that may be more generalizable, and constructing concentration-response functions that can be used in risk assessments.

The largest multi-city time-series study that has been conducted in the US is the National Morbidity, Mortality and Air Pollution Study (NMMAPS) (Samet *et al.*, 2000). 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 (Bell *et al.*, 2007; Bell and Dominici, 2008; Smith *et al.*, 2009; Stylianou and Nicholich, 2009). The results have been assessed in the context of city-specific, regional, and national impacts of air pollution on health (*e.g.*, Figure 3.1). 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 multi-city studies, several reviews and meta-analyses of air pollution epidemiology studies have been conducted, primarily of time-series mortality studies (*e.g.*, Schwartz, 1994; Stieb *et al.*, 2002, 2003; Dominici *et al.*, 2003; Bell *et al.*, 2005; Ito *et al.*, 2005; Levy *et al.*, 2005). Some of these studies evaluated the health impacts of several air pollutants

---

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

(*e.g.*, PM_{10}, SO_2, NO_2, CO, and O_3), both independently and in the multi-pollutant context (*e.g.*, Stieb *et al*. 2002). Others have focused on only one specific pollutant (*e.g.*, O_3) but considered the potential impacts of confounding co-pollutants (*e.g.*, PM) (*e.g.*, Bell *et al*., 2005, Levy *et al*., 2005). Overall, these reviews have explored the role of confounding factors in exposure-outcome relationships, identified systematic approaches for selecting time-series studies and extracting data for meta-analysis, and evaluated the relative merits of various statistical models used to analyze time-series data (*i.e.*, the GLM and GAM) and the degree to which these models accurately 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 effects [*e.g.*, neurocognitive effects (Kaufman, 2001), behavioral effects (Goodlad *et al*., 2013), cardiovascular effects such as blood pressure impacts (Navas-Acien *et al*., 2008), and cancer (Fu and Boffetta, 1995)], potential effect markers [*e.g.*, genetic polymorphisms for δ-aminolevulinic acid dehydratase, an enzyme involved in heme biosynthesis (ALAD); Scinicariello *et al*., 2007], and exposure issues such as effectiveness of exposure reduction interventions (Yeoh *et al*., 2012). These analyses have yielded estimates of the associations between lead exposure and various outcomes, odds or risk ratios, and concentration- or dose-response relationships.

Overall, meta-analyses (and multi-city studies) of observational epidemiology studies are but one component of US EPA's descriptive discussions of specific health endpoints in US EPA's NAAQS evaluations. For example, for lead, meta-analyses have been included in discussions of findings regarding blood pressure (Navas-Acien *et al*., 2008) and conduct disorder (Marcus *et al*., 2010). For several other criteria air pollutants (*e.g.*, PM and O_3), multi-city studies have played a fundamental role in policy decisions regarding the NAAQS. The effort to date to address the challenges of the diverse air pollution health effects literature in the context of NAAQS development provides opportunities to learn from previous applications of meta-analysis techniques and identify ways that this technique could be used more systematically and effectively in the future.

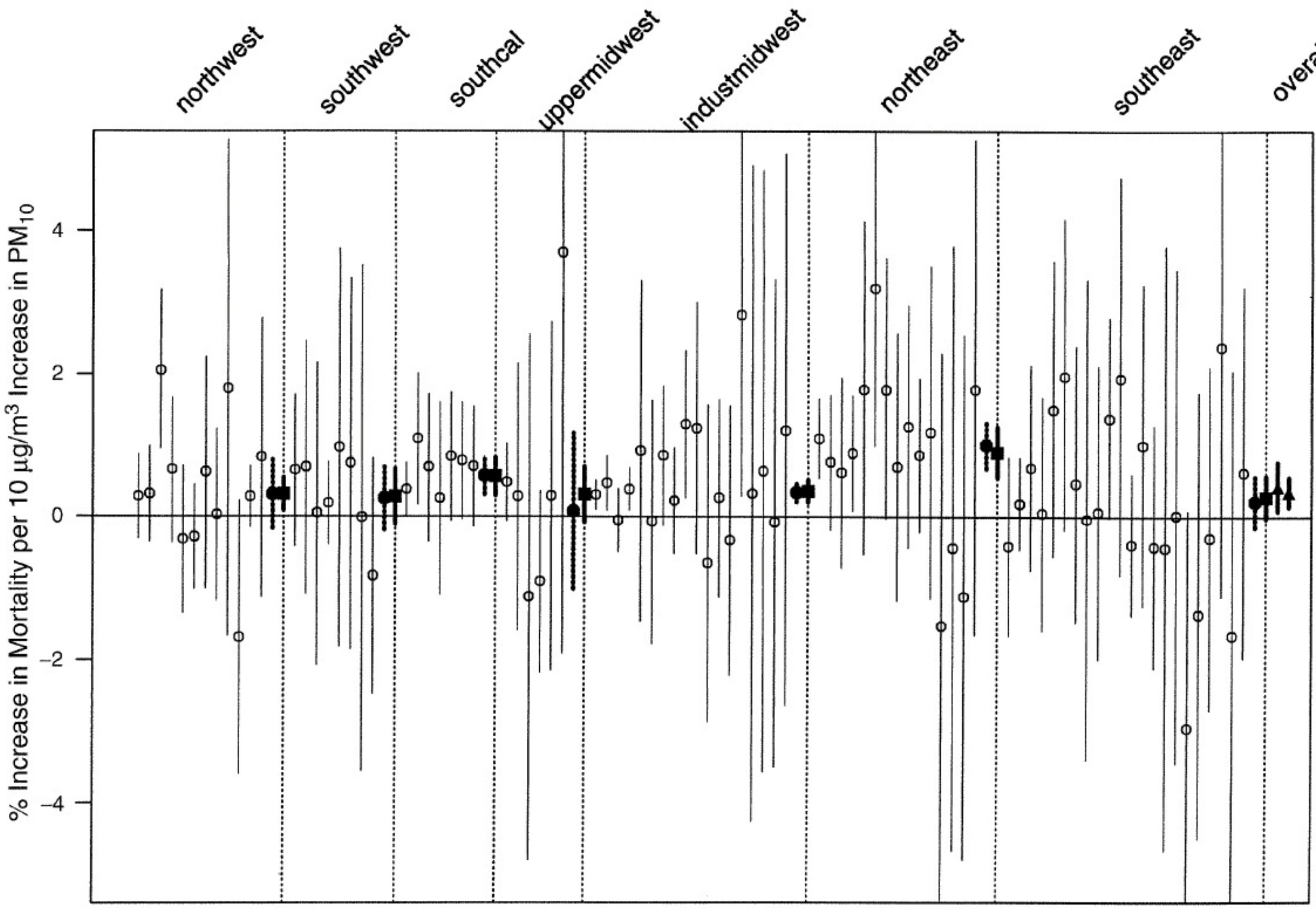

Figure 3.1 Percent Increase in Total Mortality Associated with a 10 μg/m^3 Increase in PM_{10} in 90 NMMAPS Cities, with 95% Confidence Intervals and Grouped by Region. This figure illustrates another approach to displaying meta-analysis results using a forest plot. The open circles represent specific 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*. (2000).

3.2 Meta-analysis Interpretation Issues

Like the meta-analyses conducted using NO_2 controlled exposure studies, the observational epidemiology literature for criteria air pollutant substances contains examples of how differences in meta-analysis choices and approaches can yield different results and interpretations. As illustrated below, two key issues to consider when interpreting the findings from meta-analyses are bias and heterogeneity.

3.2.1 Bias

The applications of meta-analysis in air pollutant research reflect the importance of careful consideration of potential bias in study data collection, presentation, analysis, and synthesis. For example, the meta-analyses of O_3 time-series studies illustrate how bias can be introduced into data evaluations by the choice of averaging time or lag period between exposure and effect that is used in analyzing and reporting the study results (*i.e.*, selection bias). Analyses have shown that reported effect estimates can be biased if researchers choose to report time-series study results based on the averaging time or lag time that is most significant across studies, rather than choosing to analyze the data based on a single consistent averaging time or lag time for all studies. For example, both Levy *et al*. (2005) and Bell *et al*. (2005) found higher effect estimates when using results based on the most statistically significant lag times reported in the individual studies *vs*. using a consistent lag time of 0 in the effects analyses. In a meta-analysis of asthma incidence and long-term air pollution exposures, Anderson *et al*. (2013) reported publication bias, as well as bias associated with conducting evaluations that are not systematic in selecting and extracting study data (*i.e.*, using only the effect estimates that were statistically significant in their meta-analysis *versus* using those that were selected based on *a priori* selection criteria).

Meta-analyses of observational epidemiology literature regarding air pollutants have also provided insights into a potential source of bias in results from single-city studies relative to publications addressing results from coordinated multi-city studies. For example, as shown in Figure 3.2, Bell *et al*. (2005) observed that when they compared the results from their meta-analyses of single-city studies to those from the NMMAPS summary multi-city estimates, their meta-analyses yielded significantly higher mortality estimates. The authors suggested that this finding may be due to publication bias, *i.e.*, the NMMAPS reports of findings from multiple cities may more routinely include negative results from individual cities, while researchers conducting single city studies with negative findings may be less likely to submit such findings for publication.

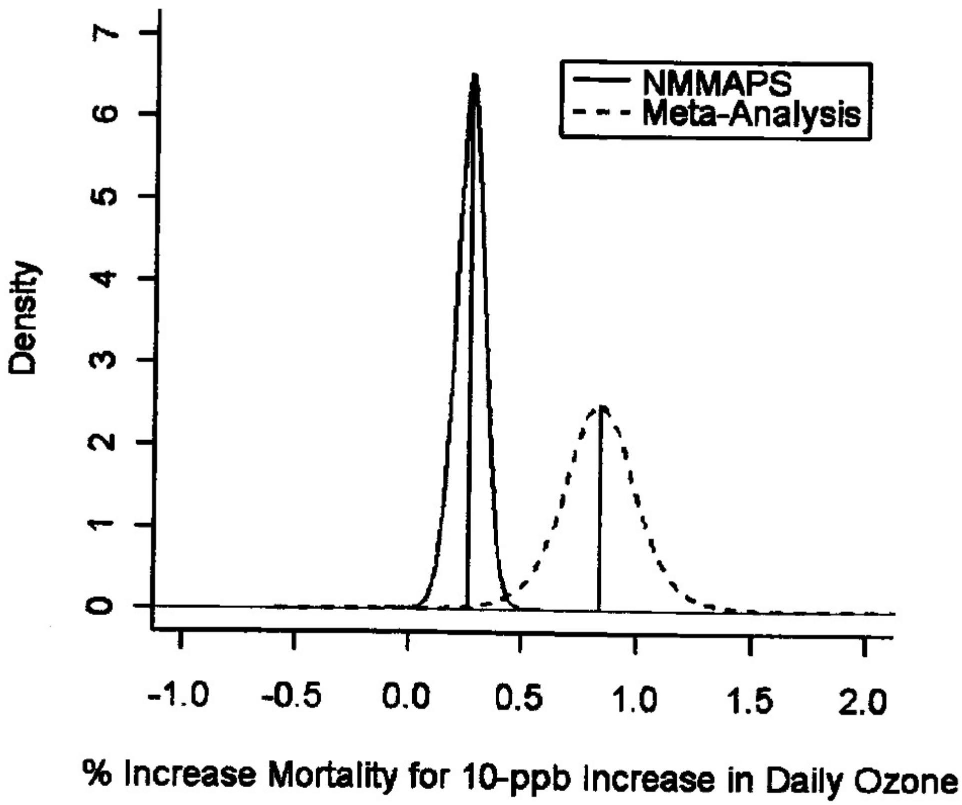

Figure 3.2 Distributions of Summary Log-Relative Risks of All-cause Mortality Associated with a 10-ppb Increase in O_3 in 95 Cities (NMMAPS) Compared to a Meta-analysis of 11 US Estimates. The multi-city results yielded a lower and more precise estimate of the overall percent decrease in mortality associated with O_3 exposures than did the meta-analysis based on studies reporting results from single cities. Source: Bell *et al*. (2005).

The issue of potential impacts of bias on interpretation of results also arises in studies of lead. For example, Bellinger (2009) identified numerous potential opportunities for selective reporting or reporting bias in neurotoxicity studies due to the many choices that researchers can make regarding which 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 (2009) also presented several illustrative analyses demonstrating how conclusions drawn from data 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 determine covariate inclusion in the analyses (Stewart, 2006, as cited in Bellinger, 2009).

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 in the analysis (Meyer-Baron and Seeber, 2000; Goodman *et al*., 2001, 2002; Schwartz *et al*., 2002; Seeber *et al*., 2002; Seeber and Meyer-Baron, 2003). Specifically, one meta-analysis of 22 studies of neurobehavioral effects in occupational populations exposed to lead (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 neurobehavioural effects of exposure to moderate blood concentrations of lead" (Goodman *et al*., 2002). Seeber *et al*. (2002) 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]." Among the areas of debate between these research groups regarding meta-analysis approaches and interpretation were 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*., 2002) 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 that may introduce bias into the meta-analysis. 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; Goodlad *et al*., 2013). As recognized by US EPA's Clean Air Science Advisory Committee in its review of US EPA's *Integrated Science Assessment for Lead* (CASAC, 2013), a substantial design limitation in most studies of potential associations between lead exposure and ADHD is failure to include or adequately assess information regarding parental psychopathology, a fundamental factor that 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*. (2013), "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 of any limitations or other notable features of the included studies that are not directly reflected in the meta-analysis approach.

3.2.2 Heterogeneity

Heterogeneity is a pervasive challenge in synthesizing results from observational epidemiology studies. As a result, air pollutant research reflects a number of efforts to enhance the consistency of certain categories of available studies and improve the ability to more effectively combine and compare study results. As described above, as one mechanism to develop more readily integrated data representing a range of locations, some researchers have conducted multi-city studies applying a common research approach at numerous sites.

Despite efforts to reduce the effects of heterogeneity on systematic evaluations of air pollution research, differences remain among studies with respect to such factors as outcome definitions, study populations (*e.g.*, age), study periods (*e.g.*, seasonal *versus* 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 particular, issues related to the use of the GAM model were identified that suggested this model overestimated effects for some air pollutants (Stieb *et al*., 2003). For O_3, additional issues arise because air monitoring data are limited to the summer season for many areas of the US. In addition, researchers have employed different exposure averaging times to evaluate O_3 effects (*e.g.*, 24-hr average, 8-hr average, and 1-hr maximum), the mortality estimates for which are not equivalent. If these differences are not addressed appropriately in data analyses (including meta-analyses), conclusions drawn may be unreliable.

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 US 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 (*e.g.*, Mazumdar *et al*., 2011). 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 (Braun *et al*., 2012). As discussed below, efforts toward enhancing 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 less consistency was apparent in the approaches used to analyze and report the study results. In particular, Thacker *et al*. (1992) 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 insufficient information was provided to allow development of 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 identified by Thacker *et al*. (1992) that hampered conducting a meta-analysis of the longitudinal lead studies included conflicting results and inconsistent patterns of regression and correlation coefficients (*i.e.*, heterogeneity). As a result, despite efforts to enhance the comparability of the studies, they were insufficient to support the more detailed comparisons and analyses of a formal meta-analysis. To support greater consistency in study reporting and collection of data in a centralized location, Thacker *et al*. (1992) 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 quantification of air standards for lead as well as other regulatory and risk assessment settings (*e.g.*, Lanphear *et al*., 2005). However, this pooled analysis reflects only a small portion of the health effects literature available for lead, and researchers have noted that studies of neurodevelopmental impacts (of lead and other substances), as well as other areas of epidemiological research would benefit from use of more consistent analytical and reporting approaches that would ease study comparisons and synthesis (*e.g.*, Bellinger, 2000, 2007, 2009). In particular, focusing on neurotoxicity data, Bellinger (2009) 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 for synthesizing data from observational epidemiology studies discussed above 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 of the scientific literature 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" (Goodman, M *et al*., 2010). 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, 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, to assist comparisons with previous studies. 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). The general observations reflected in their recommendations are clearly transferable to other research areas. Moreover, as recognized by Meyer-Baron *et al*. (2011), enhancing the ability of researchers to systematically synthesize and summarize available research findings is not only important for developing sound interpretations of available study results, but it 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 non-chemical stressors (*e.g.*, socioeconomic variables) (Levy *et al.*, 2013). Although some efforts have been made to meta-analyze multi-pollutant data, the lack of consistently reported results from multi-pollutant analyses has hindered proper data synthesis (Stieb *et al*., 2002). Opportunities remain for better data evaluation and reporting to enhance synthesis across studies of the findings from analyses with multiple pollutants or other factors.

New developments in 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 multi-city study analysis, provide opportunities for evaluating factors that contribute to heterogeneity in the context of evaluating single-city studies. Hybrid meta-analytic approaches are also being developed to incorporate uncertainty associated with combining information from a limited number of studies (*e.g.*, Shin *et al*., 2013, Levy *et al*., 2013).

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 in the examples discussed above, approaches used to report study data can influence the ability of researchers to synthesize study findings, as well as the potential for bias to be 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 multi-city 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 time have been based primarily on human data, but the causation evaluations that underlie them also consider toxicology data. Results are often available from studies examining endpoints in animals relevant to the causal questions posed by the NAAQS process and, although meta-analyses have traditionally been used mostly for human data (Glass, 1976), they can be a helpful method for synthesizing animal data for specific endpoints and determining whether those data are robust.

Animal toxicity studies can be heterogeneous as a result of the use of different species, study designs, and protocols. However, they may be more homogeneous than observational epidemiology studies owing to better control of exposures, test conditions, and outcome assessments. Thus, evaluations of data from studies of laboratory animals 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 would strengthen the meta-analyses of the resulting data. In addition, the increased precision of meta-analyses as compared to analyses of individual studies can aid in reducing 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 (Peters *et al*., 2006).

Many types of data from experimental studies using laboratory animals can be summarized and quantified using meta-analysis approaches. Data from laboratory animal studies 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*. (2006) determined that researchers most commonly used simple methods for performing a quantitative synthesis of results across studies, such as the calculation of mean or median values of outcome measures. Other methods have also been applied, such as fixed and random effects precision-weighted models and exposure-response models.

For example, Valberg and Crouch (1999) conducted a meta-analysis of data regarding lung tumors in rats following lifetime inhalation of diesel exhaust particulates (DEPs). The authors directly evaluated the raw data from eight individual studies for statistical evidence of a threshold in lung tumor response between high and low exposure concentrations. They used a multi-stage 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*. (2013) used a simpler approach based on determining the number of studies with positive findings at each dose for each specific endpoint. The authors stated that, because of the variety of animal species and endpoints measured across the different studies included in the meta-analysis, calculation of a summary estimate of effect size was not possible. They determined that the pattern of positive results for the *in vivo* toxicity of nano-titanium dioxide was dependent upon the dose, exposure route, and organ examined, and they also observed that the highest percentage of positive studies reported effects in the liver and kidney. These findings were not evident from a review of the individual studies.

Meta-analyses of animal toxicity studies can help determine the consistency and generalizability of effects of chemical exposures, but several factors must be considered. As with meta-analyses of human data, publication bias can significantly affect the interpretation of a meta-analysis of laboratory animal data, leading to overestimation of treatment-related effects. In addition, as noted above, between-study heterogeneity is a common feature of meta-analyses that must be addressed. Some of this heterogeneity comes from differences in animal species used across studies. However, studies using different species can be included together 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 taken into account in the statistical models (Peters *et al.*, 2006).

A major problem associated with meta-analyses of animal toxicity data is the large number of published studies with incomplete reporting of study design and methods. There are no widely used guidelines 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. A lack of information on a given parameter can introduce bias into the study, as well as any meta-analysis that incorporates the study. Failure to consider differences across studies in the statistical models due to this lack of information can also result in reduced statistical power and false positive results (Tseng *et al.*, 2012). If possible, all experimental variables should be considered and 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 (Peters *et al.*, 2006; Macleod *et al.*, 2009; Hooijmans *et al.*, 2010; Kilkenny *et al.*, 2010; van der Worp *et al.*, 2010). For example, Hooijmans *et al*. (2010) developed a "gold standard publication checklist" of items that should be included in every published animal study and Kilkenny *et al*. (2010) 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 includes the use of a sample size calculation prior to the start of the study. In a related effort to strengthen animal studies and the usefulness of their results, 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) (Krauth *et al*., 2013).

Although they are not currently used in causation evaluations supporting NAAQS determinations, meta-analyses of animal toxicity studies can lead to a better interpretation of existing results from primary studies, which can inform causality determinations by providing plausibility for associations observed in human studies. A meta-analysis offers a framework for investigating potential publication bias, which can lead to overestimation of treatment effects and make the evidence unreliable for regulatory decision making. Through an understanding of the sources of bias that may be apparent in laboratory animal studies, the quality of conducting and reporting these studies may be improved. Such improvements in the underlying scientific studies would contribute to regulatory decision making that is 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 (including those that generate a large amount of data) can also be amenable for use in meta-analyses. Combining data from multiple mechanistic studies in a meta-analysis can lead to a better understanding of 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 be used to understand the toxicity of chemicals at the cellular or molecular level. These 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. This technology has been evolving over the past two decades and is being used in a wide array of contexts, providing an opportunity for innovative applications of meta-analysis. 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 results in a larger data set, which increases the statistical power to obtain a more precise estimate of 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 (Pennings *et al*., 2008). There are many gene expression meta-analyses in the published literature, and Ramasamy *et al*. (2008) outlined practical guidelines for conducting a meta-analysis of microarray data sets in seven distinct steps.

There are several challenges for conducting meta-analyses of gene expression data. One is the quality of the data in terms of reporting of phenotypic information about the biological samples examined. A set of criteria called MIAME (Minimum Information About a Microarray Experiment) was developed for researchers to provide information on the necessary experimental conditions for verifying and reproducing results of microarray studies (Brazma *et al*., 2001). Microarray data submitted to public repositories, as well as to many scientific journals for publication, must be MIAME-compliant, but often there is incomplete information on 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 (Schmidberger *et al*., 2011). The inclusion of as much biological information as possible in the reporting of individual gene expression studies is necessary for the reliability and overall quality of meta-analyses that include these studies.

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 be used to alleviate this issue (Tseng *et al*., 2012).

A further challenge for combining gene expression data from multiple studies is the technical complexity of integrating data across multiple microarray platforms. There are many microarray platforms available, with overlapping sets of gene probes across platforms. While some normalization procedures require all studies in a meta-analysis of microarray data 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 a direct merge of data sets (Tseng *et al*., 2012).

The examination of gene expression changes in cells or tissues from different species can also be a source of between-study heterogeneity, as there is often large variability between gene expression patterns from different organisms. On the other hand, 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*., 2013).

In addition to gene expression and other types of *in vitro* studies, *in vivo* mechanistic data from studies in experimental animals (such as those discussed in the previous section) and humans can also be combined using meta-analysis methods. For example, Nakao *et al*. (2011) investigated whether the behavioral and cognitive deficits of attention-deficit hyperactivity disorder (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 (Brubaker *et al.*, 2009, 2010; Cecil *et al.*, 2008, 2011) 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 be 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 the NAAQS as more mechanistic studies are conducted. By providing a more synthesized interpretation of results regarding a chemical's MoA, meta-analyses of mechanistic studies can inform the understanding of whether associations from epidemiology and animal toxicity studies are biologically plausible. 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 the criteria air pollutant research examples discussed in this paper, 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 be used to 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 the ability of researchers to draw well-supported conclusions from studies.

Meta-analysis techniques can also allow researchers to draw upon a broader database for 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 (TCEQ, 2011). Specifically, instead of deriving a toxicity value based on a dose or exposure level drawn from a single study, TCEQ integrated three values from two studies of lung cancer in workers with nickel exposures to derive a final URF that reflected the relative value and significance of the data derived from each of the selected studies. 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-analysis efforts and the meta-analysis process itself can yield insights to support scientific understanding and policy decision-making. Most notably, conducting the systematic study review required for a meta-analysis can help researchers resolve and understand the basis for apparent inconsistencies among the results of individual studies. Study comparisons in a meta-analysis framework can also help identify false positives in individual studies, insights regarding specific factors influencing study results, and whether study findings can be generalized to other populations. Furthermore, sensitivity analysis within a meta-analysis framework can indicate the robustness of the available data and the degree to which the overall study findings are influenced by the results of specific studies. 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 or influencing factors). Meta-analysis tools cannot be applied in all circumstances – and cannot directly encompass all available data regarding a specific research question; the studies included in a specific 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 (Blair *et al*., 1995), 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, there is limited access to data of sufficient quality, or there is substantial variation in study design or population. In addition, important differences in effect estimates, exposure metrics, or other factors may limit or even preclude quantitative statistical combination of multiple studies [such as where different measures or test systems are used to assess the magnitude of a given health effect, or different approaches are used to define exposures, *e.g.*, using different exposure categories (quartiles/quintiles) or linear *vs*. log-transformed data]. 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 in the qualitative discussion of the results (Blair *et al*., 1995).

Moreover, meta-analyses alone cannot address the adversity of the outcome being studied. Such determinations require consideration of the degree to which the outcome measure is related to the actual adverse effect of concern or reflects functional impairment. For example, Goodman, JE *et al*. (2010) 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 paper illustrate other key issues for meta-analysis related to specific types of data and opportunities for using meta-analysis to strengthen scientific understanding and policy evaluations for air pollutants. In one specific example discussed in this paper, the analyses of the available controlled exposure data for NO_2 demonstrate the impacts that design choices can have on the results of meta-analyses, how those results are interpreted, and policy decisions that rely on those analyses. Specifically, use of a more refined approach (*e.g.*, incorporating more use of stratified evaluations and meta-regressions) offers a better understanding of the data and can help regulators avoid making policy decisions based on erroneous data interpretations. Meta-analyses that evaluated associations between short-term exposures to O_3 and mortality highlighted the need to consider publication bias, stratified analyses of seasonal effects, the choice of study estimates, and multi-pollutant evaluations. As also reflected in the examples from the air pollutant literature, advances in analyzing and interpreting the available data using meta-analysis approaches can result from more effective applications of existing tools as well as development of more sophisticated methodologies. As a general observation, numerous opportunities exist for expanding the use of meta-analysis approaches to more systematically synthesize the diverse, multi-faceted 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 factors, particularly heterogeneity in study design, influencing the ability of researchers to conduct meta-analyses. Use of more consistent and comprehensive study designs can also enhance the usefulness of small studies by providing a way to aggregate such data. Care may be needed to avoid excessive consistency, *i.e.*, the extensive use of common designs would make it difficult to determine whether or how study results might differ if certain study elements were changed.

As illustrated by the longitudinal studies of children's exposures to lead, in which efforts were made to coordinate study design and implementation with the goal of facilitating data synthesis (*e.g.*, Bornschein and Rabinowitz, 1985), the lack of attention to consistency in data analysis and reporting of results hampered efforts to synthesize the study results (Thacker *et al*., 1992). Similar observations regarding the need for greater consistency in study design and data analysis to better support data synthesis have continued to be made in the more recent scientific literature regarding this research area, *e.g.*, regarding the neurotoxicity literature for lead (*e.g.*, Bellinger, 2009), and other compounds, such as PCBs (Goodman, M *et al*., 2010). Goodman, M *et al*. (2010) 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 for comparisons and syntheses among studies conducted at different points in time.

The benefits of greater consistency among studies for efforts to synthesize study findings have been acknowledged and reflected in a number of systematic review guidelines for conducting studies or reporting results. One of the seminal efforts to promote more systematic evaluations of scientific data regarding human health issues, 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 (Higgins and Green, 2011), 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, use of this handbook could greatly improve consistency of these studies, which in turn would form a basis for comparing results. A key component of the handbook is adherence 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 analysis 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*. (1995) 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*. (2000) focused on issues for reporting results from Meta-analysis Of Observational Studies in Epidemiology (MOOSE), while Liberati *et al*. (2009) 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 the reporting of laboratory animal meta-analyses (Peters *et al*., 2006), as well as for reporting laboratory animal study data to better support meta-analyses and other systematic data reviews (*e.g.*, Macleod *et al*., 2009; Hooijmans *et al*., 2010; Kilkenny *et al*., 2010; van der Worp *et al*., 2010). 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 any sensitivity analyses that were performed. Application of these approaches would enhance interpretation, synthesis, and understanding of meta-analysis results.

The air pollutant research reviewed in this paper 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 the evaluation of causation for environmental contaminants. As such, opportunities exist to help proactively shape this research to more effectively support data syntheses using meta-analysis techniques. In other areas of research, such as animal toxicity studies, opportunities exist for better coordination of 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 (Morales *et al*., 2009), 2,4,6-trichorophenol (Xu *et al*., 2011), organophosphate pesticides

(Bouchard *et al*., 2010), perfluorinated compounds (Gump *et al*., 2011), phthalate metabolites (Engel *et al*., 2010), and a variety of compounds characterized as endocrine disruptors (de Cock *et al*., 2012). As research regarding many of these substances is in an early stage, consideration of 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, and these data could also be useful for assessing the relative magnitude of the associations of these substances with ADHD.

Studies of ecological impacts of environmental contaminants represent another area where meta-analysis could be more extensively applied. While meta-analysis has been increasingly used to evaluate ecological studies, such applications have been most prevalent in such fields as evolutionary ecology, community ecology, and conservation ecology (Gurevitch *et al*, 2001; Gates, 2002; Nakagawa and Poulin, 2012). Innovations in meta-analysis that have been explored in this research area include meta-analyses of meta-analyses (*e.g.*, in the area of plant evolutionary ecology) (Castellanos and Verdu, 2012) 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) (Nakagawa and Santos, 2012).

While meta-analysis is not the only way to systematically review research study data (*e.g.*, see Rhomberg *et al*., 2013), the rigor, specificity, and transparency of conducting and documenting this type of evaluation present particular benefits for conducting systematic data reviews and syntheses, even for situations where it may not be possible to complete a meta-analysis owing to data limitations. In particular, it encompasses a more specific, in-depth consideration of study elements that would need to be combined to conduct a meta-analysis and how those elements would need to be comparable across studies than does a more qualitative systematic review. For evaluations of specific endpoints using existing data, this detailed perspective helps in selecting studies to be included in the systematic analyses, *e.g.*, by helping to identify factors influencing study outcomes, limitations in specific studies, and appropriate study groupings to include in the meta-analyses (*e.g.*, within stratified analyses). Meta-analysis also offers options for weighting the results from studies of varying quality and strength and deriving synthesized, well-documented effect estimates reflecting those considerations. Where multi-faceted data sets need to be addressed (*e.g.*, in considering data relevant for a specific outcome across disciplines), meta-analyses can help make such data more manageable and understandable, *e.g.*, by providing 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 researcher attention on key aspects of the literature warranting additional research or evaluation in a policy-making setting. With its emphasis on quantitative syntheses, meta-analysis is particularly well suited for assessing the relative importance of endpoints and for identifying data gaps in existing knowledge (*e.g.*, for evaluating which endpoint reflects the most sensitive endpoint for policy evaluations and the degree to which such determinations are well supported). 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 (*e.g.*, Nakagawa and Santos, 2012; Gates, 2002). 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 outlook offered by meta-analysis also offers benefits for enhancing the quality of the published literature, as well as designing and implementing new research efforts. For example, greater consideration of potential applications in meta-analyses during journal publication reviews could help mitigate publication bias (for example, if the information reflected in studies with results that are negative or not statistically significant 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 be beneficial to data syntheses regardless of whether the data were applied in a quantitative meta-analysis or not.

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 types of data and its ability to aid understanding of complex data sets is particularly attractive in light of the ever-increasing amount of scientific data that is being generated and must be interpreted. 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 we have 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 paper 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 multi-faceted issue.

References

Adams, WC. 2002. "Comparison of chamber and face-mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses." *Inhal. Toxicol.* 14(7):745-764.

Adams, WC. 2006. "Comparison of chamber 6.6-h exposures to 0.04-0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses." *Inhal. Toxicol.* 18(2):127-136.

Anderson, GB; Bell, ML. 2010. "Does one size fit all? The suitability of standard ozone exposure metric conversion ratios and implications for epidemiology." *J. Expo. Sci. Environ. Epidemiol.* 20(1):2-11.

Anderson, HR; Favarato, G; Atkinson, RW. 2013. "Long-term exposure to air pollution and the incidence of asthma: Meta-analysis of cohort studies." *Air Qual. Atmos. Health* 6(1):47-56.

Bell, ML; Dominici, F. 2008. "Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities." *Am. J. Epidemiol.* 167(8):986-997.

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

Bell, ML; Kim, JY; Dominici, F. 2007. "Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies." *Environ. Health Perspect.* 115(11):1591-1595.

Bellinger, DC. 2000. "Effect modification in epidemiologic studies of low-level neurotoxicant exposures and health outcomes." *Neurotoxicol. Teratol.* 22(1):133-140.

Bellinger, DC. 2007. "Lead neurotoxicity in children: Decomposing the variability in dose-effect relationships." *Am. J. Ind. Med.* 50(10):720-728.

Bellinger, DC. 2009. "Interpreting epidemiologic studies of developmental neurotoxicity: Conceptual and analytic issues." *Neurotoxicol. Teratol.* 31(5):267-274.

Blair, A; Burg, J; Foran, J; Gibb, H; Greenland, S; Morris, R; Raabe, G; Savitz, D; Teta, J; Wartenberg, D; Wong, O; Zimmerman, R. 1995. "Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute." *Regul. Toxicol. Pharmacol.* 22(2):189-197.

Bornschein, RL; Rabinowitz, MB; Eds. 1985. "The Second International Conference on Prospective Studies of Lead, Cincinnati, Ohio, April 9-11, 1984." *Environ. Res.* 38(1).

Bouchard, MF; Bellinger, DC; Wright, RO; Weisskopf, MG. 2010. "Attention-deficit/hyperactivity disorder and urinary metabolites of organophosphate pesticides." *Pediatrics* 125(6):e1270-e1277.

Braun, JM; Hoffman, E; Schwartz, 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. 2012. "Assessing windows of susceptibility to lead-induced cognitive deficits in Mexican children." *Neurotoxicology* 33(5):1040-1047.

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. 2001. "Minimum information about a microarray experiment (MIAME) -Toward standards for microarray data." *Nat. Genet*. 29(4):365-371.

Brubaker, CJ; Dietrich, KN; Lanphear, BP; Cecil, KM. 2010. "The influence of age of lead exposure on adult gray matter volume." *Neurotoxicology* 31(3):259-266.

Brubaker, CJ; Schmithorst, VJ; Haynes, EN; Dietrich, KN; Egelhoff, JC; Lindquist, DM; Lanphear, BP; Cecil, KM. 2009. "Altered myelination and axonal integrity in adults with childhood lead exposure: A diffusion tensor imaging study." *Neurotoxicology* 30(6):867-875.

Castellanos, MC; Verdu, M. 2012. "Meta-analysis of meta-analyses in plant evolutionary ecology." *Evol. Ecol.* 26(5):1187-1196.

Cecil, KM; Brubaker, CJ; Adler, CM; Dietrich, KN; Altaye, M; Egelhoff, JC; Wessel, S; Elangovan, I; Hornung, R; Jarvis, K; Lanphear, BP. 2008. "Decreased brain volume in adults with childhood lead exposure." *PLoS Med.* 5(5):e112.

Cecil, KM; Dietrich, KN; Altaye, M; Egelhoff, JC; Lindquist, DM; Brubaker, CJ; Lanphear, BP. 2011. "Proton magnetic resonance spectroscopy in adults with childhood lead exposure." *Environ. Health Perspect.* 119(3):403-408.

Centre for Reviews and Dissemination (CRD). 2009. "Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care." Accessed on August 06, 2013 at http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf, 294p., January.

Chang, X; Zhang, Y; Tang, M; Wang, B. 2013. "Health effects of exposure to nano-TiO2: A meta-analysis of experimental studies." *Nanoscale Res. Lett.* 8(1):51.

Clean Air Science Advisory Committee (CASAC). 2013. Letter to B. Perciasepe (EPA) re: CASAC Review of EPA's "Integrated Science Assessment for Lead (Third External Review Draft - November 2012)". Science Advisory Board, EPA-CASAC-13-004. 84p., June 4.

Cockcroft, DW; Davis, BE. 2006. "Mechanisms of airway hyperresponsiveness." *J. Allergy Clin. Immunol.* 118(3):551-559.

Cockcroft, DW; Davis, BE; Boulet, LP; Deschesnes, F; Gauvreau, GM; O'Byrne, PM; Watson, RM. 2005. "The links between allergen skin test sensitivity, airway responsiveness and airway response to allergen." *Allergy* 60(1):56-59.

de Cock, M; Maas, YG; van de Bor, M. 2012. "Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review." *Acta Paediatr.* 101(8):811-818.

Dominici, F; Sheppard, L; Clyde, M. 2003. "Health effects of air pollution: A statistical review." *Int. Stat. Rev.* 71:243-276.

Egger, M; Smith, GD; Altman, D. 2001. *Systematic Reviews in Health Care: Meta-Analysis in Context (Second Edition)*. BMJ Publishing Group, London, 487p.

Engel, SM; Miodovnik, A; Canfield, RL; Zhu, C; Silva, MJ; Calafat, AM; Wolff, MS. 2010. "Prenatal phthalate exposure is associated with childhood behavior and executive functioning." *Environ. Health Perspect.* 118(4):565-571.

Fu, H; Boffetta, P. 1995. "Cancer and occupational exposure to inorganic lead compounds: a meta-analysis of published data." *Occup. Environ. Med.* 52(2):73-81.

Gates, S. 2002. "Review of methodology of quantitative reviews using meta-analysis in ecology." *J. Anim. Ecol.* 71(4):547-557.

Glass, GV. 1976. "Primary, secondary, and meta-analysis of research." *Educ. Res.* 5(10):3-8.

Goodlad, JK; Marcus, DK; Fulton, JJ. 2013. "Lead and Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms: A meta-analysis." *Clin. Psychol. Rev.* 33(3):417-425.

Goodman, JE; Chandalia, JK; Thakali, S; Seeley, M. 2009. "Meta-analysis of nitrogen dioxide exposure and airway hyper-responsiveness in asthmatics." *Crit. Rev. Toxicol.* 39(9):719-742.

Goodman, M; Squibb, K; Youngstrom, E; Anthony, LG; Kenworthy, L; Lipkin, PH; Mattison, DR; Lakind, JS. 2010. "Using systematic reviews and meta-analyses to support regulatory decision-making for neurotoxicants: Lessons learned from a case study of PCBs." *Environ. Health Perspect.* 118:727-734.

Goodman, JE; Dodge, DG; Bailey, LA. 2010. "A framework for assessing causality and adverse effects in humans with a case study of sulfur dioxide." *Regul. Toxicol. Pharmacol.* 58:308-322.

Goodman, M; LaVerda, N; Clarke, C; Foster, ED; Iannuzzi, J; Mandel, J. 2002. "Neurobehavioral testing in workers occupationally exposed to lead: Systematic review and meta-analysis of publications." *Occup. Environ. Med.* 59(4):217-223.

Goodman, M; LaVerda, N; Mandel, J. 2001. "Commentary on "A meta-analysis for neurobehavioral results due to occupational lead exposure with blood lead concentrations <70 micrograms/100 ml."" *Arch. Toxicol.* 75(7):439-440.

Gump, BB; Wu, Q; Dumas, AK; Kannan, K. 2011. "Perfluorochemical (PFC) exposure in children: Associations with impaired response inhibition." *Environ. Sci. Technol.* 45(19):8151-8159.

Gurevitch, J; Curtis, PS; Jones, MH. 2001. "Meta-analysis in ecology." *Adv. Ecol. Res.* 32:199-247.

Higgins, JPT; Green, S. 2011. "Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)." The Cochrane Collaboration. Accessed on February 03, 2012 at http://www.cochrane-handbook.org/, March.

Hooijmans, CR; Leenaars, M; Ritskes-Hoitinga, M. 2010. "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." *Altern. Lab. Anim.* 38(2):167-182.

Ito, K; De Leon, SF; Lippmann, M. 2005. "Associations between ozone and daily mortality: Analysis and meta-analysis." *Epidemiology* 16(4):446-457.

Jerrett, M; Burnett, RT; Pope, CA; Ito, K; Thurston, G; Krewski, D; Shi, Y; Calle, E; Thun, M. 2009. "Long-term ozone exposure and mortality." *N. Engl. J. Med*. 360(11):1085-1095.

Kaufman, AS. 2001. "Do low levels of lead produce IQ loss in children? A careful examination of the literature." Arch. Clin. Neuropsychol. 16(4):303-341.

Kilkenny, C; Browne, WJ; Cuthill, IC; Emerson, M; Altman DG. 2010. "Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research." *PLoS Biol.* 8(6):e1000412.

Kim, CS; Alexis, NE; Rappold, AG; Kehrl, H; Hazucha, MJ; Lay, JC; Schmitt, MT; Case, M; Devlin, RB; Peden, DB; Diaz-Sanchez, D. 2011. "Lung function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 hours." *Am. J. Respir. Crit. Care Med.* 183:1215-1221.

Krauth, D; Woodruff, TJ; Bero, L. 2013. "Instruments for assessing risk of bias and other methodological criteria of published animal studies: A systematic review." *Environ. Health Perspect*. 121(9):985-992.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, A; Thurston, G; Calle, EE; Thun, MJ. 2009. "Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." Health Effects Institute (HEI) Research Report 140, Boston, MA. 154p., May.

Kristiansson, E; Osterlund, T; Gunnarsson, L; Arne, G; Joakim Larsson, DG; Nerman, O. 2013. "A novel method for cross-species gene expression analysis." *BMC Bioinformatics* doi:10.1186/1471-2105-14-70.

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. 2005. "Low-level environmental lead exposure and children's intellectual function: An international pooled analysis." *Environ. Health Perspect.* 113(7):894-899.

Levy, JI; Fabian, MP; Peters, JL. 2013. "Meta-Analytic Approaches for Multi-Stressor Dose-Response Function Development: Strengths, Limitations, and Case Studies." 20p., August.

Levy, JI; Chemerynski, SM; Sarnat, JA. 2005. "Ozone exposure and mortality: An empiric Bayes metaregression analysis." *Epidemiology* 16 (4) :458-468.

Liberati, A; Altman, DG; Tetzlaff, J; Mulrow, C; Gatzsche, PC; Ioannidis, JPA; Clarke, M; Devereaux, PJ; Kleijnen, J; Moher, D. 2009. "The PRISMA Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care intverventions: Explanation and elaboration." *Ann. Intern. Med.* 151:W65-W94.

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. 2009. "Reprint: Good laboratory practice: Preventing introduction of bias at the bench." *J. Cereb. Blood Flow Metab.* 29(2):221-223.

Marcus, DK; Fulton, JJ; Clarke, EJ. 2010. "Lead and conduct problems: A meta-analysis." *J. Clin. Child Adolesc. Psychol.* 39(2):234-241.

Mazumdar, M; Bellinger, DC; Gregas, M; Abanilla, K; Bacic, J; Needleman, HL. 2011. "Low-level environmental lead exposure in childhood and adult intellectual function: A follow-up study." *Environ. Health* 10:24.

McDonnell, WF. 1993. "Utility of controlled human exposure studies for assessing the health effects of complex mixtures and indoor air pollutants." *Environ. Health Perspect.* 101(Suppl. 4):199-203.

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. 2011. "Statistical means to enhance the comparability of data within a pooled analysis of individual data in neurobehavioral toxicology." *Toxicol. Lett.* 206(2):144-151.

Meyer-Baron, M; Seeber, A. 2000. "A meta-analysis for neurobehavioral results due to occupational lead exposure with blood lead concentrations <70 microg/100 ml." *Arch. Toxicol.* 73(10/11):510-518.

Morales, E; Julvez, J; Torrent, M; de Cid, R; Guxens, M; Bustamante, M; Kunzli, N; Sunyer, J. 2009. "Association of early-life exposure to household gas appliances and indoor nitrogen dioxide with cognition and attention behavior in preschoolers." *Am. J. Epidemiol.* 169(11):1327-1336.

Nakagawa, S; Poulin, R. 2012. "Meta-analytic insights into evolutionary ecology: An introduction and synthesis." *Evol. Ecol.* 26(5):1085-1099.

Nakagawa, S; Santos, ESA. 2012. "Methodological issues and advances in biological meta-analysis." *Evol. Ecol.* 26(5):1253-1274.

Nakao, T; Radua, J; Rubia, K; Mataix-Cols, D. 2011. "Gray matter volume abnormalities in ADHD: Voxel-based meta-analysis exploring the effects of age and stimulant medication." *Am. J. Psychiatry* 168(11):1154-1163.

Navas-Acien, A; Schwartz, BS; Rothenberg, SJ; Hu, H; Silbergeld, EK; Guallar, E. 2008. "Bone lead levels and blood pressure endpoints: A meta-analysis." *Epidemiology* 19(3):496-504.

Nordmann, AJ; Kasenda, B; Briel, M. 2012. "Meta-analyses: What they can and cannot do." *Swiss Med. Wkly.* 142:w13518.

Pennings, JL; Kimman, TG; Janssen, R. 2008. "Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques." *PLoS ONE* 3(7):e2596.

Peters, JL; Sutton, AJ; Jones, DR; Rushton, L; Abrams, KR. 2006. "A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting." *J. Environ. Sci. Health B* 41(7):1245-1258.

Ramasamy, A; Mondry, A; Holmes, CC; Altman, DG. 2008. "Key issues in conducting a meta-analysis of gene expression microarray datasets." *PLoS Med.* 5(9):e184.

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. 2013. "A survey of frameworks for best practices in weight-of-evidence analyses." *Crit. Rev. Tox.* 43(9):753-784.

Samet, JM; Zeger, SI; Dominici, F; et al. June 2000. "The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity, mortality, and air pollution in the United States." Health Effects Institute (Cambridge, MA) HEI Research Report 94, Part II. 84p.

Schelegle, ES; Morales, CA; Walby, WF; Marion, S; Allen, RP. 2009. "6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans." *Am. J. Respir. Crit. Care Med.* 180(3):265-272.

Schmidberger, M; Lennert, S; Mansmann, U. 2011. "Conceptual aspects of large meta-analyses with publicly available microarray data: A case study in oncology." *Bioinform. Biol. Insights* 5:13-39.

Schwartz, BS; Stewart, W; Hu, H. 2002. "Neurobehavioral testing in workers occupationally exposed to lead (Letter)." *Occup. Environ. Med.* 59:648-650.

Schwartz, J. 1994. "Air pollution and daily mortality: A review and meta analysis." *Environ. Res.* 64:36-52.

Scinicariello, F; Murray, HE; Moffett, DB; Abadin, HG; Sexton, MJ; Fowler, BA. 2007. "Lead and delta-aminolevulinic acid dehydratase polymorphism: Where does it lead? A meta-analysis." *Environ. Health Perspect.* 115(1):35-41.

Seeber, A; Meyer-Baron, M. 2003. "Neurobehavioral testing in workers occupationally exposed to lead (Letter and authors' reply)." *Occup. Environ. Med.* 60:145-146.

Seeber, A; Meyer-Baron, M; Schaper, M. 2002. "A summary of two meta-analyses on neurobehavioral effects due to occupational lead exposure." *Arch. Toxicol.* 76(3):137-145.

Shin, HH; Cohen, A; Pope, CA III; Ezzati, M; Lim, SS; Hubbell, B; Burnett, RT. 2013. "Critical Issues in Combining Disparate Sources of Information to Estimate the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure." 24p., August.

Smith, RL; Xu, B; Switzer, P. 2009. "Reassessing the relationship between ozone and short-term mortality in U.S. urban communities." *Inhal. Toxicol.* 21 (Suppl. 2) :37-61.

Stieb, DM; Judek, S; Burnett, RT. 2002. "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." *J. Air Waste Manage. Assoc.* 52 :470-484.

Stieb, DM; Judek, S; Burnett, RT. 2003. "Meta-analysis of time-series studies of air pollution and mortality: Update in relation to the use of generalized additive models." *J. Air Waste Manage. Assoc.* 53(3):258-261.

Stroup, DF; Berlin, JA; Morton, SC; Olkin, I; Williamson, GD; Rennie, D; Moher, D; Becker, BJ; Sipe, TA; Thacker, SB. 2000. "Meta-analysis of observational studies in epidemiology: A proposal for reporting." *JAMA* 283:2008-2012.

Stylianou, M; Nicolich, MJ. 2009. "Cumulative effects and threshold levels in air pollution mortality: Data analysis of nine large US cities using the NMMAPS dataset." *Environ. Pollut.* 157(8-9):2216-2223.

Texas Commission on Environmental Quality (TCEQ). 2011. "Nickel and Inorganic Nickel Compounds (Final)." Toxicology Division. Accessed on August 15, 2013 at http://www.tceq.state.tx.us/assets/public/implementation/tox/dsd/final/june11/nickel_&_compounds.pdf, 130p., June 1.

Thacker, SB; Hoffman, DA; Smith, J; Steinberg, K; Zack, M. 1992. "Effect of low-level body burdens of lead on the mental development of children: Limitations of meta-analysis in a review of longitudinal data." *Arch. Environ. Health* 47(5):336-346.

Tseng, GC; Ghosh, D; Feingold, E. 2012. "Comprehensive literature review and statistical considerations for microarray meta-analysis." *Nucleic Acids Res*. 40(9):3785-3799.

US EPA. 2007. "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. 609p., July.

US EPA. 2008. "Integrated Science Assessment for Oxides of Nitrogen." National Center for Environmental Assessment (NCEA) - RTP Division, EPA/600/R-08/071. 260p., July.

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. 2228p., December.

US EPA. 2009b. "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. 896p., July.

US EPA. 2010a. "Integrated Science Assessment for Carbon Monoxide." National Center for Environmental Assessment (NCEA) - RTP Division, EPA/600/R-09/019F. 593p., January.

US EPA. 2010b. "Primary national ambient air quality standards for nitrogen dioxide (Final rule)."*Fed. Reg.* 75(26):6473-6538. February 9.

US EPA. 2012. "Clean Air Act [website]." Accessed August 6, 2013 at http://www.epa.gov/air/caa/, 1p., February 17.

US EPA. 2013a. "Integrated Science Assessment for Lead (Final)." National Center for Environmental Assessment (NCEA), EPA/600/R-10/075F. 1885p., June.

US EPA. 2013b. "Integrated Science Assessment for Ozone and Related Photochemical Oxidants (Final)." National Center for Environmental Assessment (NCEA), EPA/600/R–10/076F. 1251p., February.

US EPA. 2013c. "Nitrogen Dioxide (NO_2) Primary Standards." Accessed August 6, 2013 at http://www.epa.gov/ttn/naaqs/standards/nox/s_nox_index.html, 1p., July 12.

US EPA. 2013d. "Sulfur Dioxide (SO_2) Primary National Ambient Air Quality Standards." Accessed August 6, 2013 at http://www.epa.gov/ttn/naaqs/standards/so2/s_so2_index.html, 1p., July 12.

Valberg, PA; Crouch, EAC. 1999. "Meta-analysis of rat lung tumors from lifetime inhalation of diesel exhaust." *Environ. Health Perspect.* 107(9):693-699.

van der Worp, HB; Howells, DW; Sena, ES; Porritt, MJ; Rewell, S; O'Collins, V; Macleod, MR. 2010. "Can animal models of disease reliably inform human studies?" *PLoS Med.* 7(3):e1000245.

Xu, X; Nembhard, WN; Kan, H; Kearney, G; Zhang, ZJ; Talbott, EO. 2011. "Urinary trichlorophenol levels and increased risk of attention deficit hyperactivity disorder among US school-aged children." *Occup. Environ. Med.* 68(8):557-561.

Yeoh, B; Woolfenden, S; Lanphear, B; Ridley, GF; Livingstone, N. 2012. "Household interventions for preventing domestic lead exposure in children (Review)." *Cochrane Database Syst. Rev.* 4:CD006047.