Meta-Analysis:
Recycling Garbage or an Important Tool for Evaluating the Evidence?

Bruce R. Parker

Venable LLP

750 E. Pratt Street, Suite 900
Baltimore, MD 21202
(410) 244-7534

Dominik D. Alexander

EpidStat Institute

2100 Commonwealth Blvd., Suite 203
Ann Arbor, MI 48105
(303) 670-6734
Bruce R. Parker, a partner with Venable LLP, has served on several mass tort litigation trial teams and has tried to verdict cases involving Actos, Mirapex, latex gloves, and breast implants. A fellow in the ACTL, he has also served as lead counsel in numerous MDL Daubert hearings. He has served as president of the IADC and the Maryland Defense Counsel, director of the IADC Trial Academy, dean for the IADC Corporate Counsel College, and on the DRI Board of Directors and the Board of Directors of Lawyers for Civil Justice.

Dominik D. Alexander, PhD, MSPH, with the EpidStat Institute, has extensive experience in health research methodology, meta-analysis, and disease causation, particularly in the conceptualization, design, analysis, and interpretation of epidemiologic studies. He has published on a diverse range of topics and types of studies, including original epidemiologic research, qualitative reviews, systematic weight-of-evidence assessments, and quantitative meta-analyses. He has over 150 peer-reviewed manuscripts, professional presentations, abstracts, and book chapters. Dr. Alexander was awarded a National Cancer Institute Fellowship for Cancer Prevention and Control.

The authors wish to acknowledge the assistance of James K. O'Connor in performing research support.
Meta-Analysis: Recycling Garbage or an Important Tool for Evaluating the Evidence?

Table of Contents

I. Introduction ...................................................................................................................................................5
II. Meta-Analysis Overview ...............................................................................................................................5
III. Background and Relevance of Meta-Analysis ..............................................................................................5
IV. Utilization of Meta-Analysis Methodology ..................................................................................................6
    A. Precision .................................................................................................................................................. 6
    B. Consistency .............................................................................................................................................7
    C. Heterogeneity .......................................................................................................................................... 8
    E. Publication Bias ......................................................................................................................................9
    F. Evidence Synthesis ...............................................................................................................................10
    G. Objectivity (Meta) v. Subjectivity (Narrative) ....................................................................................11
    H. A Meta-Analysis Inherently Examines Study Quality .......................................................................11
    I. General Steps for Conducting a Meta-Analysis ..................................................................................12
V. Practice Pointers on Dealing with Meta-Analysis .....................................................................................13
    A. Attacking Meta-Analysis-Based Expert Opinion ...............................................................................13
    B. Affirmative Use of Meta-Analysis-Based Expert Opinion .................................................................15
VI. Summary ......................................................................................................................................................16
VII. Reference List ...............................................................................................................................................16
Meta-Analysis: Recycling Garbage or an Important Tool for Evaluating the Evidence?

I. Introduction

Meta-analysis is a statistical tool that, like any tool found in a hardware store, can be very helpful when used in the right manner, but, when misused, can make the job more difficult, or even damage it. Meta-analysis is a method for synthesizing data across a body of evidence to estimate statistical associations and to determine whether the associations represent “true” risk. It should come as no surprise to any defense lawyer that plaintiffs’ experts misuse this tool to create associations that don’t exist. The difficulty for the defense lawyer is being able to demonstrate in an understandable manner to a jury that corners have been cut by the expert performing the meta-analysis and how doing so produced a false result.

The goal of this paper is to teach lawyers what meta-analysis is and how it is performed correctly. We discuss case law on the admissibility of expert testimony that relies upon meta-analysis and we offer some practice pointers on cross-examining a plaintiff’s expert who has improperly used meta-analytic techniques to evaluate a body of evidence.

II. Meta-Analysis Overview

Interpreting epidemiologic and scientific studies of various exposures and disease outcomes is complex because it involves many methodological issues and sources of uncertainty. Such uncertainty makes reliable scientific interpretation an even greater challenge for scientists, public health officials, regulators, and litigators. While evaluation of each individual study is of fundamental importance when reviewing a body of studies, systematically conducting a comprehensive meta-analysis of data across studies can greatly augment interpretation of the evidence base. Furthermore, a meta-analytic approach can supersede the review of studies on an individual basis by synthesizing and summarizing the collective body of evidence in a transparent and uniform fashion. Although a meta-analysis synthesizes data across all studies, each individual study entry in a meta-analysis contributes weight and influence to the total body of evidence. A meta-analysis can be used to: generate an estimate of effect change (or relative risk, depending on the nature of the study methodology) with significantly enhanced precision, examine consistency of associations by study characteristics, identify sources of between-study variation (heterogeneity), discern potential patterns of associations by study factors (e.g., dose levels, changes in effect over time, differences by gender or age), and assess the potential for reporting or publication bias. Thus, a meta-analysis can serve as an integral component in a weight-of-evidence assessment.

An exponential increase in the utilization of meta-analysis methodology in virtually all scientific disciplines has been adopted over the past decade, owing to its value in evaluating and interpreting scientific evidence. As a result, there has been a monotonic rise in published meta-analyses in peer-reviewed medical and scientific journals around the world. However, the quality of the published meta-analyses is variable. Unfortunately, a non-trivial proportion of published meta-analyses convolute interpretation rather than make the scientific evidence clearer. Thus, the following sections summarize the usefulness and validity of meta-analysis methodology, how meta-analyses can be used objectively and appropriately, and areas where meta-analyses may go wrong, and how to identify such areas.

III. Background and Relevance of Meta-Analysis

Meta-analysis is a systematic method used to evaluate quantitatively a body of literature. Meta-analyses are often considered the gold standard for evidence-based reviews when applied to high quality stud-
ies. Numerous top tier journals have adopted scoring systems and checklists for conducting meta-analyses. Furthermore, many governing agencies and committees, such as the Dietary Guidelines for Americans, the Environmental Protection Agency (EPA), and the International Agency for Research on Cancer (IARC), use meta-analyses to summarize a body of evidence for which recommendations are made. In the U.S. Food and Drug Administration (FDA) guidance document entitled, *Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims*, it is stated that, the FDA “…intends to consider as part of its health claim review process a meta-analysis that reviews all the publicly available studies on the substance/disease relationship”. How well such agencies conduct and interpret their quantitative meta-analyses may be a matter of debate. This discussion is beyond the scope of this paper.

A meta-analysis is a systematic statistical approach to combining the results from individual studies in order to generate a single comprehensive statistic, commonly referred to as a summary relative risk estimate (SRRE) or a weighted group mean difference (WGMD) (Alexander et al. 2016; Alexander et al. 2015; Walker et al. 2008). A quantitative technique in and of itself, a meta-analysis is able to integrate results from both qualitative and quantitative studies (Deeks et al. 2011). A meta-analysis is designed to be a comprehensive measure of all available primary research for a particular exposure and outcome (Egger and Smith 1997). As the volume of studies published increases, more data become available that facilitate formal meta-analysis methods for systematically synthesizing data. Although traditional narrative reviews and meta-analyses both seek to summarize the current state-of-the-science for a specific research topic by evaluating all available relevant literature, a meta-analysis differs from a narrative review by following standard guidelines and procedures to review data from individual studies and importantly, they are designed to calculate summary measures of effect.

Meta-analyses can serve many fundamental components (*e.g.*, systematic identification of the relevant literature, quantitative summary of the evidence, evaluation of study quality) in a weight-of-evidence assessment. Their primary capabilities are used to: estimate effect sizes (*e.g.*, hazards ratios of survival comparing treatments, group mean changes in body composition) with significantly enhanced precision (*i.e.*, increasing the power to detect a difference by combining data across studies) when performed correctly, examine consistency of associations overall and by specific study characteristics, identify sources of between-study variation (heterogeneity), discern potential patterns of associations by study factors (*e.g.*, dose levels, changes in effect over time, differences by gender or age), and assess the potential for reporting or publication bias.

**IV. Utilization of Meta-Analysis Methodology**

**A. Precision**

A paramount goal of a meta-analysis is to provide a more precise estimate of the direction and magnitude of effect of an intervention or exposure (Deeks et al. 2011). Traditionally, meta-analysis has been used to arrive at summary conclusions about a body of research when individual studies are too small to generate valid conclusions (Petitti 2000). A synthesis (meta-analysis of all available studies) can help researchers place the results of one study in the context of other similar studies (Borenstein et al. 2009; Greenland and O’Rourke 2008). It can also be used to discern differences among specific effect estimates of many individual studies, which greatly enhances the utility compared to evaluating the evidence on a study by study basis. Thus, a meta-analysis can overcome issues in interpretation that arise from reviewing individual studies of small sample sizes, including smaller sub-group comparisons (Alexander et al. 2016; Walker et al. 2008). For example, it may be difficult to interpret a study or group of studies that have small sample sizes where the analyses yield associations with very wide confidence intervals. By combining such studies meta-analytically, the statistical power is increased and the analysis yields a more precise estimate (*i.e.*, narrower confidence intervals) of effect
that may be more interpretable. This also holds true when combining data for a smaller sub-group of studies for a particular factor (e.g., trial duration) or a refined parameter within an individual study (e.g., analyses by dose). The enhanced precision may reveal a statistically significant difference between sub-groups – something that may go undetected when reviewing the literature on a study by study basis. By extension, a meta-analysis can demonstrate minor or small effects among large sample sizes, which are not often achieved by individual studies (Egger et al. 2001). For example, to detect a 10% reduction in risk of myocardial infarction with the use of a particular drug with greater than 90% certainty, a randomized controlled trial may require at least ten thousand subjects (Collins et al. 1992). A meta-analysis can avoid the costly burden of performing this study by summarizing the effects of many smaller trials (Egger et al. 2001; Noble, Jr. 2006).

By virtue of conducting a meta-analysis, the ‘best’ estimate of the effect is generated. This occurs because in a meta-analysis model, each individual study provides weight (i.e., statistical influence) that contributes to the overall effect. This is considered the totality of the quantitative evidence, and each study plays a role in contributing to the evidence. Thus, the technique of a meta-analysis, in and of itself, is a quantitative summary of the totality of the evidence that produces the best estimate of effect with enhanced statistical precision.

Although there are obvious benefits to having a more precise estimate of risk, meta-analyses with unsound methodology will produce summary effects that may be misleading or biased. Thus, there may be a perceived level of precision that is not substantiated by the underlying data sources. An objective and transparent protocol should be followed and documented so the reader can understand fully, the process of study identification and data inclusion in an analytical model. It is encouraged that the reader interprets the summary effect and confidence intervals based on the methods for study inclusion in order to validate the analytical results.

B. Consistency

A meta-analysis can provide important insights into many aspects of causation and can serve an important role when evaluating the totality of the data. A primary strength of meta-analysis is its capacity to better characterize the existence and nature of associations summarized across studies. Meta-analysis can be used to identify patterns among results of many studies, sources of discrepancies among those results, or other interesting relationships that may come to light in the context of many other studies (Alexander et al. 2016; Greenland and O’Rourke 2008). A meta-analysis provides an assessment of the consistency of associations and may facilitate the identification of potential sources of between-study variation, and thus, may settle controversies between variations in directions or magnitude of effects between individual studies (Deeks et al. 2011). Specifically, a meta-analysis allows for the creation of sub-groups to evaluate the consistency or patterns of associations by relevant factors. For example, if the overall effect size (i.e., summary association from a meta-analysis) of a collection of studies is null (or heterogeneous), sub-groups can be created by certain factors (e.g., dose level, study design) that may reveal consistent patterns of associations. This important type of empirical analysis can only be performed using meta-analysis methodology. The quantitative demonstration of patterns of associations across subgroups is difficult to do when reviewing individual studies.

Consistency is an important factor in Sir Bradford Hill’s considerations for causality. Consistency is not examined at the individual study level. Rather, consistency must be based upon the critical review of a body of studies. The design of a meta-analysis allows for such a review – all studies are systematically collected and evaluated quantitatively, and each study provides influence, or statistical weight to the overall effect size. However, a meta-analysis may give a false sense of consistency if the researchers bias their selection and inclusion of studies in the analyses. Again, an objective and transparent data extraction protocol should be followed and documented so the reader can follow the methodological approach to arriving at the summary result.
C. Heterogeneity

Another important function of a meta-analysis is testing for heterogeneity. Heterogeneity reflects unexplained variation between study results in a meta-analysis model, and a model that has significant heterogeneity may indicate the need for further analytical exploration. Thus, consistency is related to heterogeneity, or in other words, if significant heterogeneity is present in a meta-analysis, the results are inconsistent. Comprehensive examinations of heterogeneity are crucial components of a meta-analysis and are necessary to fully explore the level of consistency in a body of literature while identifying factors that may contribute to varying patterns of results between studies. Statistical heterogeneity is typically tested using the Cochran’s Q test, which indicates whether the variance between studies differs, and the $I^2$ statistic, which indicates the percentage of variation attributable to between-study heterogeneity (Higgins 2002). When a significant amount of statistical heterogeneity exists, p-value for the Q test will be low and the $I^2$ statistic will be high.

Between-study variation may be the result of differences in study design, measurement techniques, patterns of associations by gender or dose level, surgical techniques, or a plethora of other factors. If there is significant heterogeneity, the results of the meta-analysis may not be generalizable as a whole or clinically reliable due to a lack of integrative research (Walker et al. 2008; Deeks et al. 2011; Borenstein et al. 2009). However, the presence of heterogeneity between studies can be perceived as an opportunity to further investigate and develop hypotheses for the differences between studies (Finckh and Tramer 2008). Indeed, meta-analyses are commonly used to identify sources of variation by conducting sub-group analyses to distinguish any differences in results by specific study characteristics. For example, the following figure is a forest plot illustration of a meta-analysis on a dietary supplement and risk of prostate cancer. The point estimates are denoted by the boxes, with the study sizes represented by the size of the boxes. The confidence intervals are represented by the horizontal lines and if the lines cross the null value of 1.0, that means there is a non-statistically significant result. Overall, there appears to be a non-statistically significant summary association that is below 1.0. In other words, a non-significant decrease in risk, with statistically significant heterogeneity ($p < 0.05$). Upon closer inspection, it is apparent that most of the larger studies show a statistically significant increase in risk while the mid- to smaller sized studies show a statistically significant decrease in risk. Thus, given this heterogeneity, a quantitative summarization of these data do not appear warranted. However, this provides an opportunity to ‘dig deeper’ into the data and identify the possible reasons for this variation. In this case, the statistical discrepancy was due to the way exposure was ascertained; the studies in the positive direction were based on self-reported intake whereas the studies in the inverse direction were based on biological measurements of intake levels. In summary, by looking more closely at the data, it may be possible to discern patterns of associations and reasons why and how individual study results differ.

There are two commonly used models in meta-analysis; fixed-effects and random-effects. When using a fixed-effects model, it is assumed that there is one ‘true’ effect size for all studies in the analysis, and that all results between studies are due to sampling error. However, it is rare to ascribe all variation to sampling error, as the overwhelming majority of meta-analyses combine data from different study populations, from different researchers, from different geographic locations, and from different points in time among other factors. Thus, a random-effects model allows for the fact that the true effect size may differ from study to study. A random-effects model accounts for both, within-study as well as between-study variation. In contrast, a fixed-effects model accounts for within-study variation. Therefore, there is an extra calculation for between-study variability in the algorithm for random-effects models. In general, this additional variation component results in a more conservative distribution of statistical weight given to each of the individual studies in a meta-analysis model, and thus, the confidence intervals in a random-effects model are appropriately wider than fixed-effects models. In some cases, the sampling errors may be minor and the sample sizes of the studies may be very large, resulting in a convergence of the random-effects model to the fixed-effects model. This occurs because the between-study variation component in the random-effects algorithm is nearing zero, leaving a single component (within-study variation) that is present in both random and fixed-effects models.

It is commonly suggested that if statistical heterogeneity is present, a random-effects model should always be used. However, given the often considerable variability between studies in terms of population characteristics, exposure and outcome assessments, and study designs, a random effects model should always be advocated regardless of the presence of heterogeneity or the perceived absence of heterogeneity. In general, it is easier to achieve statistical significance when using fixed-effects models because the confidence intervals are narrower (or precision is enhanced). Thus, it is recommended that readers should interpret the results based on the model selection used in a meta-analysis. If a statistically significant effect is reported, and the only model used is fixed-effects, then the published result may not be an accurate representation of the true range of values in the context of methodological variation.

E. Publication Bias

Meta-analytically, researchers can only summarize the data that are available. There may be a tendency of scientific journals to primarily publish studies with strong positive associations or statistically significant results, and those studies that contain negative or null findings may be absent from the scientific literature. This results in publication bias. Researchers may also withhold null or non-statistically significant findings from their manuscript. This is considered reporting bias, which is a type of publication bias. When reviewing studies on an individual basis, the potential impact or relevance of publication bias cannot be assessed. However, when using meta-analysis methodology, publication bias can be assessed both visually (typically by using a funnel plot) and statistically (such as Egger’s regression test) (Walker et al. 2008; Egger et al. 2001; Finckh and Tramer 2008).

Funnel plots show the distribution of individual study effect sizes against a measure of study size, which may include the standard error of the measure, such as a hazard ratio. When there is no evidence of publication bias, the funnel plot will be a symmetrical display of study data around the summary effect size. However, when publication bias is indicated the scatter plots of study data will be asymmetrical. Asymmetry is commonly observed when there is a greater predilection of smaller studies with results on the positive side of the group estimate to be published in a selective fashion. Other methods to detect potential publication bias are statistically driven. For example, when there is funnel plot asymmetry, a regression method (referred to as the Egger test) can be conducted with a two-sided p-value to determine whether there is a statistically
significant deviation from symmetry. Ideally, a larger number of studies (e.g., 10 or more) may be necessary to achieve the desired statistical power of regression testing for publication bias. Other, less common methods include the Failsafe N method and the “Trim and Fill” method. The Failsafe N method was originally adopted by social scientists to determine how many studies (that are presumably null) it would take to reduce the overall effect size from statistically significant to non-statistically significant. In theory, this is a scientifically attractive approach; however, the premise is that all unpublished studies are not significant and/or null. A somewhat related technique is the Trim and Fill method, whereby the studies that are asymmetric around the summary estimate are ‘trimmed’ off and imputed studies are ‘filled’ back in to create funnel plot symmetry. This method creates a theoretical symmetrical environment with the assumption that asymmetrical studies have a missing counterpart. This method, as with the Failsafe N method, should be viewed as a sensitivity analysis, rather than a statistical adjustment for the effects of publication bias.

While there are many ways to evaluate the potential for publication bias as mentioned above, one of the techniques to determine whether there is an indicator for between-study variability should be utilized. Testing for publication bias cannot immediately reveal the source of the variation between studies or why a particular study is an outlier. Publication bias tests serve as instruments to explore the reasons for the outlier values. For example, nine studies may have relative risk estimates of 0.90 to 1.10, suggesting no overall effect. However, the tenth study in the model may have a relative risk estimated of 2.30. Clearly, the tenth study is a statistical outlier. Testing for publication bias would identify this study as an outlier but it does not tell the researcher (or reader) why this study is an outlier other than the fact that the magnitude of risk is considerably larger than all other studies. Therefore, this finding is an indicator, and suggests that the researchers explore why such a strong association was observed in this study. Upon closer inspection, there may be myriad reasons why this study was different. It may be that a high-risk study population was evaluated, or that the exposure or outcome under study was classified differently than the other studies, or that important confounders were not adjusted for, or alternatively, the result may have been due to statistical variation.

F. Evidence Synthesis

Meta-analytic techniques are important for understanding the existing primary research literature on virtually all scientific topic areas (note: meta-analyses are also commonly used in other disciplines, such as economics and cognitive psychology). The number of existing studies on a particular subject varies widely and often, and the results of individual studies in and of themselves may remain unclear and limit translation when attempting to make a well-informed decision. The ability to compare and contrast results from several separate studies on a topic of interest is useful to reduce such ambiguity. The importance of a meta-analysis on evidence synthesis is not just a matter of combining separate studies to produce a singular outcome. Rather, a meta-analysis can be used to examine critically the state-of-the-science of topic areas for which a growing number of studies are published that have a varying and complex methodological framework. For example, the authors of a neurology study stated that, “Making sense of a neuroimaging literature that is growing in scope and complexity will require increasingly sophisticated tools for synthesizing findings across studies. Meta-analysis of neuroimaging studies fills a unique niche in this process: It can be used to evaluate the consistency of findings across different laboratories and task variants, and it can be used to evaluate the specificity of findings in brain regions or networks to particular task types” (Wager et al. 2009). Thus, the researchers used meta-analysis methodology to understand better a complex topic area while focusing on consistency and specificity to make informed decisions on a body of literature. Therefore, meta-analyses are capable of summarizing available data and identifying patterns, disagreement, and/or relationships that may exist among the data that could otherwise remain unrealized (Walker et al. 2008).
Of increasing public health and clinical importance, a meta-analysis can reduce the possibility of false negatives (or false positives) in the results, which are difficult to exclude in individual studies (Finckh and Tramer 2008; Noble, Jr. 2006; Zwahlen et al. 2008). Small sample sizes of existing studies may undermine the reliability of produced results, yielding what may be considered a false positive or negative – an artifactual finding that does not reflect the true evidence base. Even among studies conducted in identical populations, effect estimates will inevitably vary due to sampling variability. Therefore, a single study typically cannot identify, or exclude with certainty, small effects between exposures or treatments (Egger et al. 2001; Zwahlen et al. 2008; Egger and Smith 1997). Through the pooling of data from multiple studies, a meta-analysis can dramatically increase the sample size used to produce the overall statistic, giving the result greater statistical power (see above section on Precision). Additionally, the pooling of populations allows for a more diverse study population with greater generalizability (that is, a meta-analysis may reduce the potential for selection bias). Individual studies can only investigate a specific type of exposure or intervention, while a meta-analysis has the statistical power to evaluate several research hypotheses (Deeks et al. 2011). While meta-analysis is a powerful tool that can reduce ambiguity, provide stronger conclusions, and better inform the application of results to evidence-based decision-making, if poorly conducted, a meta-analyses may yield a false sense of consistency in the literature.

G. Objectivity (Meta) v. Subjectivity (Narrative)

Reviewing studies on an individual basis (only) and summarizing the evidence in such a manner is commonly considered a narrative exercise. When doing so, a systematic and a priori protocol is often not in place. This may lead to a subjective interpretation of the evidence which does not take into account the quantitative impact of confounding and bias. In contrast, a meta-analysis provides a more rigorous and comprehensible evaluation of past research studies than the traditional narrative review (Greenland and O’Rourke 2008). Meta-analyses (a quantitative systematic review) aim to be more objective than traditional narrative reviews by following standard guidelines and procedures. Formal and standardized rules for study identification and study inclusion criteria may minimize disagreement or obfuscation from reviewers on issues such as which studies to include or exclude in the analysis. Meta-analyses allow for the evaluation of random errors from individual studies, a phenomenon that cannot be appreciated in narrative reviews (Egger et al. 2001; Egger et al. 2001; Zwahlen et al. 2008; Egger and Smith 1997). Furthermore, the standardized process of meta-analysis should provide more transparency in how researchers arrive at their conclusions than the often murky results and conclusions of narrative reviews (Egger and Smith 1997).

H. A Meta-Analysis Inherently Examines Study Quality

The value and utility of a meta-analysis is largely dependent upon the type of information on which it is based, the clarity of methodology and reporting, the quality and comprehensiveness of the systematic process, and the interpretation of the literature. It is important to consider the methodological quality of studies that are included in a meta-analysis since the results of a meta-analysis are only as valid as the studies included in the model. This has been referred to as the “garbage in, garbage out” phenomenon. Therefore, when conducting a meta-analysis, the studies of lower methodological quality (or those not meeting the objective and transparent inclusion criteria) may not fulfill the systematic protocol to be included in the analysis. Importantly, this type of ‘process’ should be transparent and objective, and the design and conduct of meta-analyses should follow standardized guidelines for quality (Weed 1997; Moher et al. 2009; Stroup et al. 2000). If the quality of the studies included in the review are compromised and/or prone to biases, a synthesis of their results will not be able to eliminate these original flaws. For example, if a meta-analysis is conducted
on a group of case-control studies for which selection bias and measurement error may have been substantial limitations, producing a summary association from these studies is not helpful. On the other hand, a meta-analysis of well-conducted randomized controlled clinical trials may produce an accurate and valid summary association and allow for the evaluation of patterns of associations across population sub-groups (Borenstein et al. 2009; Petitti 2000; Egger et al. 2001; Egger et al. 2001; Noble, Jr. 2006; Zwahlen et al. 2008).

Importantly, in a meta-analysis, analytical models can and should be based on study characteristics, including quality factors. For example, separate models can be created that include only cohort studies or case-control studies. Or, on the other hand, the meta-analysis inclusion criteria may be that studies are required to be a specific design. Moreover, models can be created to examine the associations among studies that controlled for smoking, or studies of sufficient follow-up duration to observe an effect, or studies that were conducted in the United States vs. other countries. The point is, by design, a meta-analysis inherently is constructed to evaluate important factors that may influence study results. This is a data-driven method for which quantitative differences can be generated. Thus, it is of paramount importance that meta-analyses utilize their full analytical potential, and researchers should always perform several sub-group and sensitivity analyses. One type of sensitivity analysis is the one-study-removed technique where a meta-analysis model is generated to produce a summary estimate and then subsequent analyses of this model are run by removing a single study at each iteration. This informs as to the robustness of the overall model by the influence of any single study, and at the same time, informs as to the sensitivity of the model and the quantitative impact of each study on the overall summary estimate.

I. General Steps for Conducting a Meta-Analysis

There are well-utilized guidelines for performing meta-analyses, and many peer-reviewed journals now require that all meta-analyses follow these guidelines. In fact, many journals require that prospective authors submit a flow chart diagram of the study selection process and a checklist for quality parameters, such as taking into account various biases and confounding factors. Two of the more recognized guidelines are the Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement (Moher et al. 2009) and Meta-analysis Of Observational Studies in Epidemiology (MOOSE) (Stroup et al. 2000).

In general, meta-analyses are conducted by the following steps:

1. Establish a research question.
2. Conduct a systematic literature search with documented search string terminology.
3. Filter and synthesize studies using a priori and transparent inclusion and exclusion criteria. In addition, the scientific justification for including and/or excluding certain studies should be documented.
4. A data extraction protocol should be developed and relevant qualitative information and quantitative data should be abstracted from each study into a spreadsheet.
5. The overall and primary analytical models should be created and meta-analyses using random-effects methods should be conducted.
6. Secondary meta-analyses, such as sub-group analyses, sensitivity analyses, and publication bias assessments should be performed. These are often the most informative analyses because they may identify any underlying patterns of associations.
7. Findings should be interpreted objectively and the conclusion should be supported by the underlying analyses, which includes a thoughtful and careful assessment of the potential impact of varying biases and confounding factors.
Finally, a report and/or manuscript should be written that transparently documents the design and analytical methodology.

V. Practice Pointers on Dealing with Meta-Analysis

Defense counsel may find themselves confronted with meta-analytics in one of four ways.

1. Plaintiffs experts use meta-analysis to claim that there is a statistically significant increase in risk. A statistically significant summary association in a meta-analysis does not mean that a causal relationship exists. Causation requires additional methodological steps, of which meta-analysis can help inform.

2. Defense experts use meta-analysis to argue that the totality of the evidence fails to show that a statistically significant increased risk.

3. Defense experts use meta-analysis to show that the data relied upon by the plaintiffs’ experts are not consistent in their finding, thus failing to meet a key Bradford Hill factor.

4. Defense or plaintiff experts may use meta-analysis to determine exposure levels for which disease risks may occur.

A. Attacking Meta-Analysis-Based Expert Opinion

The courts’ handling of challenges by defendants to the admissibility of plaintiffs’ causation testimony based on meta-analysis has been inconsistent. In In re Neurontin Mktg., Sales Practices, & Products Liab. Litig., 612 F. Supp. 2d 116 (D. Mass. 2009), the court denied the defendant’s motion for summary judgment in which the defendant had argued that general causation could not be shown. To show general causation, the plaintiffs had relied on an FDA study to support their claim that Neurontin caused suicidality. Id. at 134. The FDA had performed a meta-analysis of “199 placebo-controlled clinical studies covering 11 AEDs including Neurontin.” Id. The defendant’s epidemiology expert opined that the FDA study was methodologically flawed because it ignored significant “trial heterogeneity.” Id. at 137-38. More specifically, “sixty-one percent of all of the suicidality events … came from the lamotrigine and topiramate studies, even though these two drugs account for only thirty-eight percent of the total data.” Id. Furthermore, “these two drugs were the only ones in the analysis which demonstrated a statistically significant increased risk.” Id. The defendant’s expert observed that “[y]ou cannot come up with a reliable conclusion about the pooled risk when you have that level of heterogeneity.” Id. at 138. Additionally, the defendant’s expert “attack[ed] the FDA study for employing an odds ratio methodology … [that] excluded all studies where no suicidality events occurred.” Id. Because “only three of the forty-nine studies on Neurontin submitted to the FDA … included qualifying incidents of suicidality,” he maintained this approach “arguably skew[ed] the gabapentin-specific analysis.” Id.

The court denied the defendant’s motion holding that the “criticism of the FDAs statistical methods and conclusions … affect[ed] the weight that should be given to the study, not its admissibility.” Id. Several factors were offered by the court as relevant to its decision. First “the underlying data … were placebo-controlled, clinical studies, the ‘gold standard’ for epidemiological evidence.” Id. Second, the FDA performed two sensitivity analyses to assess the issue of trial heterogeneity and based on the results decided that “pooling the results” was justified. Id. Third, the FDA Advisory Committee voted that the increased risk applied to all 11 drugs even after considering arguments similar to those offered here by the defendant. Id. at 140. None of these factors appears to address the defense challenge to the methodology used by the FDA. It is either flawed or not. The fact that the FDA performed the meta-analysis should not play a role in determining if the methodology used was appropriate.
A different result was reached in Baker v. Chevron USA, Inc., 680 F. Supp. 2d 865 (S.D. Ohio 2010) aff’d sub nom. Baker v. Chevron U.S.A. Inc., 533 F. App’x 509 (6th Cir. 2013). In this case the plaintiffs brought suit for injuries allegedly resulting from benzene exposure. Id. The defendant’s Daubert challenge was successful with the court holding that the “Infante Meta-analysis” was inadmissible, because the underlying studies were not an appropriate fit for the facts of the case. Id. at 883-85. Specifically, the overall exposure of the “subject cohorts” was much greater than the exposure of any of the plaintiffs. Thus, it was “too great an analytical leap for the Infante Meta-analysis to provide a reliable basis for [the plaintiff’s expert’s] … causation opinions.” Id.

In In re Avandia Mktg., Sales Practices & Products Liab. Litig., 2011 WL 13576 (E.D. Pa., Jan. 4, 2011), the defendant sought to exclude the general causation opinions of three experts arguing that the underlying meta-analysis was unreliable. More specifically, plaintiffs’ experts failed to provide detailed explanations of their review criteria to justify their exclusion of studies which found no association between the defendant’s drug and the plaintiffs’ alleged harm. Id. at *3-4. The court denied the defendant’s motion explaining that the motion failed to identify any “specific flaws or limitations in the design or implementation” of the meta-analyses relied on by the plaintiffs. Id. at *4. The court noted that the plaintiffs’ experts, “consulted an extensive body of epidemiological research to support their conclusions, and evaluated and weighed the quality and usefulness of the various studies.” Id. at *15. The court felt that the plaintiffs’ experts provided reasonable explanations for why they excluded certain studies from their analysis and once again noted that “[d]ifferences in conclusions go to the weight of the evidence, and not to its admissibility. Id.

In In re Baycol Products Litig., 532 F. Supp. 2d 1029 (D. Minn. 2007), the defendants’ offered several specific criticisms of the plaintiff’s expert’s meta-analysis as a basis for their challenge to the admissibility of his general causation opinions premised on his meta-analytic data. In granting the motion, the court noted that the plaintiff’s expert, “used data from the FDA U.S. AERS, and the FDA Worldwide AERS, and did not attempt to learn if there was any overlap in the data provided.” Id. at 1040. Second, although the “AERs do not use standard definitions for terms,” plaintiff’s expert, Dr. Farquhar, took the key terms at “face value, without corroborating the information contained therein.” Id. Third, he did not account for “the ‘new drug’ effect in which AERs are made more frequently during the first years a medicine is on the market. Id. Finally, “the number used for the numerator was the AERs, but the number used for the denominator came from the National Prescription Audit Plus database. This database did not account for the many samples that were provided of the new drug. Id. at 1040-41.

Defendants can also succeed in challenging general causation opinions that are based on “cherry picked” subsets of the data that reflects an increased relative risk when a properly performed meta-analysis of the methodologically sound data set would not reflect an increased risk. For example, in In re Accutane Litigation, 2015 WL 753674 at *30 (Feb. 20, 2015) the court excluded Dr. Madigan’s causation opinion and criticized him for not performing a meta-analysis. Id. (Court noted that “Rather than conducting a meta-analysis himself of all the risk assessment studies, and possibly getting ‘closer to the truth,’ [Madigan] chose to disregard eight of [the studies]. In doing so, he ignored the knowledge learned from studying approximately 2,100,000 subjects. Instead, he relies upon a study comprised of 509 people.”).

These cases illustrate that defense motions to exclude expert testimony based on a meta-analysis should focus on attacking the inadequate methodological process by which the data were selected by the expert and/or the quality of the underlying data sets (i.e., sufficient facts or data). The Baker and Baycol Products opinions exemplify successful implementations of this approach. In contrast, the Neurontin court admitted the expert testimony despite seemingly similar valid criticisms. One possible distinguishing factor is that the Neurontin meta-analysis was performed by the FDA and approved by an FDA advisory panel. Although neither fact affects the validity of the meta-analysis, the court may have perceived this as additional indicia of reliability.
B. Affirmative Use of Meta-Analysis-Based Expert Opinion

With respect to the affirmative use of meta-analysis by defense experts, it has been the experience of one of the authors that meta-analysis is rarely used by defense experts to demonstrate the absence of a statistical association between the challenged product and the outcome of interest. The reason probably has to do with the definition of relative risk (“RR”) given to jurors. Jurors are told that a RR greater than 1 reflects a positive association while less than 1 mean a negative or protective relationship between the product and outcome. Rarely, however, is a chemical or drug shown to protect against an unexpected adverse event that is the subject of litigation. Consequently, meta-analyses used in litigation will almost never produce a relative risk less than 1 to a degree of statistical significance. It is more often the case that a defense expert using meta-analysis will show a very small (less than 1.5) statistically significant RR. Most epidemiologists uninvolved in litigation would agree that a RR less than 1.5 is not compelling evidence of an association because of the imprecision of observational epidemiologic research and the existence of unidentified confounding and bias that affects the precision of such studies. Thus, despite the “positive” result produced by a meta-analysis and the defense expert’s ability to testify that this is at best weak evidence to support the plaintiff’s case, many defense lawyers would prefer not to have their experts generate this data believing that it reduces likelihood of a successful Daubert challenge on general causation. Whether this view is accurate is beyond the scope of this paper.

Rather than using meta-analysis to generate a more precise RR, meta-analysis is more likely to be used by defense attorneys and their experts to demonstrate that the plaintiffs’ evidence lacks consistency. This can be accomplished by demonstrating statistical heterogeneity or design heterogeneity. If the goal is to demonstrate the unreliability of the plaintiffs’ meta-analysis, defense counsel may want to use empirical data suggesting the unreliability of meta-analysis compared to randomized clinical studies. For example, a paper published in the NEJM in 1997 (Discrepancies Between Meta-Analysis and Subsequent Large Randomized, Controlled Trials, 337 NEJM, 536) compared 19 meta-analyses published on different health issues before a large randomized study had been conducted on the question. For 40 primary and secondary outcomes predicted by the meta-analyses, there was only “fair” agreement between the meta-analyses and the “gold standard” RCT. The authors concluded that had no RCT been conducted, meta-analysis would have suggested treatment in 32% of cases that was not found efficacious by a RCT and a rejection of efficacious treatment in 33% of cases.

Defense counsel face a difficult task when plaintiffs’ experts have used meta-analysis that generates a statistically significant elevated relative risk. Explaining to a jury how a complex tool like meta-analysis was used inappropriately is a difficult task. Counsel needs to work with their jury consultants and graphic artists to draw upon metaphors or visual analogies that provide convincing arguments to a jury that the tool was used improperly. For example, using the term “garbage in, garbage out” can be depicted as seen below and is understandable to jurors.

![Garbage in, Garbage out](image1)

![Mixing apples and oranges](image2)
Using the metaphor, “mixing apples and oranges” is a similarly understood concept. Other images that might work might be the analogy of cooking something with ingredients that have gone bad or are of poor quality. The point is combining larger numbers of poor quality ingredients is very unlikely to produce a better finished product.

Another point that defense counsel should consider when cross-examining the plaintiffs’ expert is the demonstration of publication bias. This is difficult but if it can be shown, it is effective because it demonstrates that the bias towards publishing “negative studies” by publishers made it a foregone conclusion that the results of the meta-analysis would be positive. Thus, the results offered by the plaintiff’s experts are rigged. Sports analogies are often an effective way to teach difficult concepts. For example, assume four players who started on a major league baseball team had poor seasons with all hitting less than .170 when they went on injured reserve after 80% of the season had been completed. Computing the team batting average for the starters who ended the season would artificially increase the batting average assuming the four replacements hit more than .170 in the 20% of the season in which they started.

VI. Summary

Meta-analysis has shown itself to be a very powerful methodological tool when used properly. Defense counsel can have their experts use this tool to show that the plaintiff’s evidence that arguably supports their general causation opinion is inconsistent, biased, and/or confounded. In more limited instances, this tool can be used to show that the full data set does not establish a statistically significant elevated risk. Challenging a plaintiffs’ expert who has improperly used meta-analysis to argue that an increased risk exists is difficult. Finding an effective approach requires trial counsel working with the defense causation expert to develop a substantive attack and a graphic artist to help develop demonstrative exhibits that will help the jury to understand how the tool was used inappropriately.

VII. Reference List


