

## LEADING ARTICLE

# Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related

Douglas L Weed

Interpreting observational epidemiological evidence can involve both the quantitative method of meta-analysis and the qualitative criteria-based method of causal inference. The relationships between these two methods are examined in terms of the capacity of meta-analysis to contribute to causal claims, with special emphasis on the most commonly used causal criteria: consistency, strength of association, dose-response, and plausibility. Although meta-analysis alone is not sufficient for making causal claims, it can provide a reproducible weighted average of the estimate of effect that seems better than the rules-of-thumb (e.g. majority rules and all-or-none) often used to assess consistency. A finding of statistical heterogeneity, however, need not preclude a conclusion of consistency (e.g. consistently greater than 1.0). For the criteria of strength of association and dose-response, meta-analysis provides more precise estimates, but the causal relevance of these estimates remains a matter of judgement. Finally, meta-analysis may be used to summarize evidence from biological, clinical, and social levels of knowledge, but combining evidence across levels is beyond its current capacity. Meta-analysis has a real but limited role in causal inference, adding to an understanding of some causal criteria. Meta-analysis may also point to sources of confounding or bias in its assessment of heterogeneity.

**Keywords** Causation, epidemiology, inference, meta-analysis, systematic reviews

**Accepted** 29 November 1999

The interpretation of observational epidemiological studies offers both promise and peril.<sup>1,2</sup> Aetiological hypotheses with public health implications may gain support from these studies which in turn are subject to many alternative interpretations, especially bias and confounding. Traditionally, epidemiologists have used a mostly qualitative narrative review technique that includes causal criteria—strength, consistency, plausibility, dose-response and others—as well as considerations such as bias, confounding, and study designs.<sup>3–9</sup>

How meta-analysis can help solve this difficult interpretative problem is not immediately obvious. In a recent paper in this Journal, for example, traditional narrative reviews and meta-analyses are portrayed as methodological alternatives, each with

strengths and weaknesses.<sup>10</sup> Others argue that a systematic review may include a meta-analysis but that it should not be a prominent component of that review.<sup>11</sup>

A key unexamined concern is the relationship between meta-analysis and criteria-based methods of causal inference. This paper describes how these two methods—one more quantitative and the other more qualitative—can be used together in causal assessments.

## Background: Practical and Theoretical Accounts

The relationship between meta-analysis and causal inference has both practical and theoretical dimensions. In practice, causal claims are sometimes made in a single review in which both methods are applied to the evidence. A recent and controversial example involved induced abortion and breast cancer;<sup>12</sup> the

National Cancer Institute, Division of Cancer Prevention, EPS T-41, 6130 Executive Blvd. MSC 7105, Bethesda, Maryland, 20892–7105, USA. E-mail: dw102i@nih.gov

authors claimed that several thousand breast cancer deaths each year can be attributed to induced abortion. Another review of the same evidence—*sans* meta-analysis and published almost simultaneously—concluded that there was no association between induced abortion and breast cancer much less a causal one.<sup>13</sup>

Indeed, causal claims may appear in published narrative reviews in which meta-analysis is not performed; these claims arise from the use of methods with the traditional causal criteria at their core.<sup>14</sup> Finally, some causal claims appear to arise from meta-analysis alone with minimal reference to causal inference methods and in the absence of a systematic narrative review.<sup>15</sup>

Such a wide range of practices provides no firm basis upon which the relationship between meta-analytical and causal inference methods can be teased out. Turning to theoretical accounts, however, is only marginally helpful. Many commentators have cautioned that meta-analysis may not yet be ready for widespread use, especially for observational studies.<sup>16–21</sup> Beyond that, the main concern of methodologists is less about the role of meta-analysis in assessing causation than it is about the synthetic approach to meta-analysis that emphasizes summarization of evidence over the search for heterogeneity.<sup>22–27</sup> On the topic of causation, for example, a recent paper notes only that statistical methods alone cannot provide a causal explanation of associations.<sup>23</sup> A recent text argues that meta-analysis of observational data does not provide good evidence of causation because competing hypotheses (e.g. bias and confounding) cannot be ruled out.<sup>28</sup>

From these accounts, it is reasonable to conclude only that meta-analysis may not, by itself, provide enough information to warrant a causal claim. Although this conclusion calls into question one practice described earlier, it does not provide much help in determining the extent to which meta-analysis makes causal claims easier or more difficult. A look at the finer structure of causal inference methods, the individual causal criteria, provides clues.

## Meta-analysis and Criteria of Causation

### Consistency

One of the most frequently used criteria in the practice of causal inference<sup>14</sup> is consistency, the extent to which the association is observed in different circumstances, by different investigators, using different study designs, and in different locations.<sup>6</sup> Consistency is the epidemiologist's expanded equivalent of the experimentalist's search for replicability,<sup>29</sup> although for observational studies, replication of findings in the face of different, rather than similar, circumstances is more highly valued.<sup>8,30,31</sup>

Without meta-analysis, reviewers often assess consistency with simple quantitative summarizing techniques, tallying up the per cent of studies that are positive and then utilizing a rule-of-thumb for declaring them consistent: a simple majority in some cases, a higher threshold in others.<sup>32</sup> Indeed, it has been shown that in some controversial, high-profile situations, different rules-of-thumb for consistency (coupled with its relatively high prioritization among causal considerations) have contributed to vastly different, even exactly opposite, causal judgements.<sup>33</sup> To examine whether meta-analysis can provide a less subjective assessment of consistency requires a look at how summarization of effect estimates across studies occurs; assessing heterogeneity is a key consideration.

Heterogeneity tests assess the extent to which the studies are similar enough to warrant summarization. When a lack of statistical heterogeneity across all studies occurs, a weighted average of the estimated effect size (under one or another statistical model) seems a defensible and reproducible way to approach the assessment of consistency and therefore likely better than the rules-of-thumb used in the practice of causal inference.

When heterogeneity exists, however, then exploring its sources<sup>11</sup> will not only guard against the concerns that meta-analysis too often buries important inconsistencies in statistical aggregations<sup>19</sup> but will also provide relevant information for the practice of causal inference. For example, summarization of results within heterogeneous groups provides estimates of the overall effect by measurement technique, study design, population studied, or other sources of the observed heterogeneity, although the extent to which each heterogeneous group estimate represents bias (or not) may be a matter, like so much else in causal inference, of judgement.

Heterogeneity, however, does not necessarily preclude a conclusion of consistency in causal inference. Consider a hypothetical situation in which meta-analysis reveals heterogeneous groups of summarizable studies each of which (i.e. each group) has a statistically significant estimate in the same direction although of different magnitudes. A more extreme example can be imagined in which no known covariates are associated with variation in the study results, yet the results of all individual studies show strong effect sizes (and are heterogeneous). In both instances, it is too simplistic a rule to declare that evidence of heterogeneity disallows a causal judgement due to a lack of consistency.<sup>23</sup> Indeed, in these instances, it may be reasonable to argue that the results are both consistent (i.e. consistently greater than 1.0) *and* heterogeneous.

### Biological plausibility

The extent to which an observed association in epidemiological studies is supported or not by what is known about the mechanism of action and the underlying disease process is commonly referred to as biological plausibility or sometimes coherence.<sup>34</sup> In metaphorical terms, assessing this causal criterion is like opening a window into the many-tiered structure of biological knowledge pertinent to the aetiology of disease: cellular and subcellular systems, experimental animal models, physiological parameters and genetic or other biomarkers.<sup>35</sup> Inasmuch as several similar studies may examine the same biological parameters, meta-analysis could assist in summarizing information and in identifying heterogeneity analogous to the way in which it is used for randomized clinical trials and for epidemiological studies.<sup>36</sup> A recent example involved the role of MDR1/gp170 expression in breast cancer tumour samples wherein substantial heterogeneity was linked to differences in assay techniques.<sup>37</sup> This uncommon application of meta-analysis is a step towards a more systematic approach to the assessment of biological evidence, although it seems unlikely that meta-analysis in its current quantitative form will be useful for summarizing different kinds of studies from different levels of biological knowledge.

### Strength of association and dose-response

There is an important distinction to be made between improving the precision of the relative risk estimate—which meta-analysis offers—and interpreting the causal relevance of the

absolute value of the summarized estimate—which it does not.<sup>11,18,20,38</sup> Simply put, the practitioner of causal inference, having completed a meta-analysis, remains faced with the problem of judging whether the summarized estimate—small (i.e. weak) or large (i.e. strong)—can be explained by confounding or bias. Thus it is reasonable to accept a relative risk estimate of 2.0 emerging from a meta-analysis as a better estimate than that which we may have opined through a judicious application of our ‘judgement’, somehow estimating a summary average value from a long list of single study estimates. Nevertheless, that same conclusion, emerging as it does from a meta-analysis, provides no additional warrant for a causal claim. Put another way, a summary estimate of 2.0—whether it emerges from a meta-analysis or not—remains on the borderline of what is typically called a ‘weak’ association. The same argument applies to dose-response curves,<sup>39</sup> which can also be made more precise by judicious application of meta-analysis. Nevertheless, whatever the form of those curves—increasing, decreasing, ‘S’ or ‘J’, shaped—the practitioner of causal inference is left with the task of explaining them in terms of the biological sense of the revealed pattern, potential biases, and other concerns.

### Other criteria

Specificity, temporality, and analogy are less frequently employed criteria, but are nevertheless a part of the rich historical tradition of causal inference methods.<sup>32</sup> Meta-analysis has indirect connections to the consideration of specificity; it is possible that one group of studies will have a summarizable effect estimate that differs from another because the exposure (or disease) was measured more narrowly. Meta-analysis may also provide a way to summarize evidence that examines the relationships of time-related variables to outcomes, but it is not clear that meta-analysis has important implications for the classic consideration of the temporal order of cause and effect nor for the criterion of analogy.

### Meta-analysis, Confounding, and Bias

Causal criteria may be central to causal assessments of epidemiological evidence, but other concerns, especially confounding and bias, are also important. The extent to which meta-analysis can assist in the assessment of confounding and bias is closely related to the capacity for meta-analysis to reveal heterogeneity among studies. Potential sources of heterogeneity include study design differences, exposure assessment differences, confounders (known and unknown), and bias.<sup>11,24–26</sup> Publication bias is another concern in the practice of meta-analysis that could potentially affect interpretation of results.<sup>24</sup>

### Conclusion

Both causal inference methods and meta-analysis can appear within the context of a systematic review, the latter characterized by a clearly stated purpose, careful literature searches, explicit inclusion and exclusion criteria, assessments of study validity and thus bias, and well-articulated definitions and rules of inference for selected causal criteria.<sup>4,5</sup> Several important causal criteria may be determined or made more precise through quantitative assessments of heterogeneity and summarization of effects, in particular, using meta-analyses of groups of studies from observational or biological levels of knowledge.

Thus meta-analysis has an important role to play in causal assessments, although meta-analysis alone is not sufficient for making causal claims. Meta-analysis provides a more formal statistical approach to the criterion of consistency as well as a way to identify heterogeneous groups of studies, whether epidemiological or biological in origin. It also can provide more precise estimates of the magnitude of the effect estimate and of dose-response relationships. These are the clearest and perhaps most important links between the quantitative method of meta-analysis and the more qualitative method of causal inference.

### Discussion

These conclusions imply that practitioners of meta-analysis should refrain from making causal claims without first considering the issues brought out by applying the largely qualitative causal inference methods. Practitioners of causal inference, on the other hand, should recognise the strengths and limitations of applying meta-analysis to epidemiological evidence and understand that melding the quantitative and qualitative components of these methods within the context of a systematic review is helpful and yet also only a rough solution to a longstanding methodological problem.<sup>9</sup>

Indeed, the scope of the problem of making causal claims from scientific evidence is widening. Recently, some have argued that an understanding of the determinants and distribution of disease will be improved by examining not only epidemiological and biological evidence but also social evidence for causation.<sup>40–42</sup> Causal inference, in this expanded context, will require both an ability to combine studies within the social arena, but perhaps as importantly, to assess evidence across many levels of scientific knowledge from the molecular to the social, with classic epidemiology near the middle. Following the arguments above, meta-analysis in its current form cannot be a solution because it requires that studies be reasonably similar in design and in measurements of effect, two characteristics not shared by studies of, say, DNA repair mechanisms in tumour cell lines, tissue biomarker levels in mice exposed to an environmental insult, case-control studies of cancer in workers exposed to potential carcinogens, and analyses of the impact of income or class differences on selection processes into and out of jobs in hazardous workplaces. Indeed, making inferences across many levels of scientific knowledge may require a systems-based approach<sup>43,44</sup> not captured by the methods discussed here.

### Acknowledgements

Helpful comments and suggestions were provided by Drs Graham Colditz, Matthew Longnecker, Karin Michels, and Beverly Rockhill.

### References

- 1 Angell M. The interpretation of epidemiologic studies. *N Engl J Med* 1990;**323**:823–25.
- 2 Feinstein AR. Scientific standards in epidemiological studies of the menace of daily life. *Science* 1988;**242**:1257–63.
- 3 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;**138**:697–703.

- <sup>4</sup> Weed DL. Methodologic guidelines for review papers. *J Natl Cancer Inst* 1997;**89**:6–7.
- <sup>5</sup> Breslow RA, Ross SA, Weed DL. Quality of reviews in epidemiology. *Am J Public Health* 1998;**88**:475–77.
- <sup>6</sup> Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**58**:295–300.
- <sup>7</sup> Evans AS. *Causation and Disease: A Chronological Journey*. New York: Plenum, 1993.
- <sup>8</sup> Glynn JR. A question of attribution. *Lancet* 1993;**342**:530–32.
- <sup>9</sup> Weed DL. On the use of causal criteria. *Int J Epidemiol* 1997;**26**:1137–41.
- <sup>10</sup> Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999;**28**:1–9.
- <sup>11</sup> Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *Br Med J* 1998;**316**:140–44.
- <sup>12</sup> Brind J, Chinchilli VM, Severs WB *et al*. Induced abortion as an independent risk factor for breast cancer: a comprehensive review and meta-analysis. *J Epidemiol Community Health* 1996;**50**:481–96.
- <sup>13</sup> Michels KB, Willett WC. Does induced or spontaneous abortion affect the risk of breast cancer? *Epidemiology* 1996;**7**:521–28.
- <sup>14</sup> Weed DL, Gorelic LS. The practice of causal inference in cancer epidemiology. *Cancer Epidemiol Biomark Prevention* 1996;**5**:303–11.
- <sup>15</sup> Sood AK. Cigarette smoking and cervical cancer: meta-analysis and critical review of recent studies. *Am J Prev Med* 1991;**7**:208–13.
- <sup>16</sup> Michels KB. Quo vadis meta-analysis? A potentially dangerous tool if used without adequate rules. In: DeVita V. *Important Advances in Oncology*, Philadelphia: Lippincott, 1992, 243–48.
- <sup>17</sup> Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;**140**:771–78.
- <sup>18</sup> Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997;**50**:223–29.
- <sup>19</sup> Feinstein AR. Meta-analysis: statistical alchemy for the 21st Century. *J Clin Epidemiol* 1995;**48**:71–79.
- <sup>20</sup> Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991;**44**:127–39.
- <sup>21</sup> Bailar JC. The practice of meta-analysis. *J Clin Epidemiol* 1995;**48**:149–57.
- <sup>22</sup> Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987;**9**:1–30.
- <sup>23</sup> Greenland S. Meta-analysis. In: Rothman KJ, Greenland S (eds). *Modern Epidemiology*. 2nd Edn. Philadelphia: Lippincott-Raven, 1998, pp.643–73.
- <sup>24</sup> Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;**309**:1351–55.
- <sup>25</sup> Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *Am J Epidemiol* 1995;**142**:371–82.
- <sup>26</sup> Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995;**142**:383–87.
- <sup>27</sup> Mosteller F, Colditz GA. Understanding research synthesis. *Annu Rev Public Health* 1996;**17**:1–23.
- <sup>28</sup> Miller N, Pollock VE. Meta-analytic synthesis for theory development. In: Cooper H, Hedges LV (eds). *Handbook of Research Synthesis*. New York: Russell Sage Foundation, 1994, pp.458–83.
- <sup>29</sup> Susser M. *Causal Thinking in the Health Sciences*. New York: Oxford, 1973.
- <sup>30</sup> Schlesselman JJ. ‘Proof’ of cause and effect in epidemiologic studies: criteria for judgment. *Prev Med* 1987;**16**:195–210.
- <sup>31</sup> Elwood JM. *Causal Relationships in Medicine*. New York: Oxford, 1988.
- <sup>32</sup> Weed DL. Causal and preventive inference. In: Greenwald P, Kramer BS, Weed DL (eds). *Cancer Prevention and Control*. New York: Marcel Dekker, 1995, pp.285–302.
- <sup>33</sup> Weed DL. Underdetermination and incommensurability in contemporary epidemiology. *Kennedy Inst Ethics J* 1997;**7**:107–27.
- <sup>34</sup> Weed DL, Hursting SD. Biologic plausibility in causal inference: current method and practice. *Am J Epidemiol* 1998;**147**:415–25.
- <sup>35</sup> Defining cause. In: Tomatis L (ed.). *Cancer: Causes, Occurrence and Control*. Lyon, France: International Agency for Research on Cancer, 1990, 97–125. (IARC scientific publication no. 100.)
- <sup>36</sup> Weed DL. Meta-analysis under the microscope. *J Natl Cancer Inst* 1997;**89**:904–05.
- <sup>37</sup> Trock BJ, Leonessa F, Clarke R. Multidrug resistance in breast cancer: a meta-analysis of MDR1/gp170 expression and its possible functional significance. *J Natl Cancer Inst* 1997;**89**:917–31.
- <sup>38</sup> Dickersin K, Berlin JA. Meta-analysis: state of the science. *Epidemiol Rev* 1992;**14**:154–76.
- <sup>39</sup> Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993;**4**:218–28.
- <sup>40</sup> Susser M, Susser E. Choosing a future for epidemiology: II. from black box to Chinese boxes and ecoepidemiology. *Am J Public Health* 1996;**86**:674–77.
- <sup>41</sup> Shy CM. The failure of academic epidemiology: witness for the prosecution. *Am J Epidemiol* 1997;**145**:479–84.
- <sup>42</sup> Diez-Roux AV. On genes, individuals, society and epidemiology. *Am J Epidemiol* 1998;**148**:1027–32.
- <sup>43</sup> Weed DL. Beyond black box epidemiology. *Am J Public Health* 1998;**88**:12–14.
- <sup>44</sup> Koopman JS. Comment: emerging objectives and methods in epidemiology. *Am J Public Health* 1996;**86**:630–32.