

# Do pooled estimates from meta-analyses of observational epidemiology studies contribute to causal inference?

David A Savitz <sup>1</sup>, Francesco Forastiere<sup>2</sup>

## INTRODUCTION

The purpose of meta-analyses that generate pooled risk estimates is generally to inform causal inference, compiling evidence to address an aetiological question. Logically, drawing on information from all relevant studies is likely to be superior to relying on any single study.

Meta-analysis applies an objective, quantitative method to integrate evidence across studies. Objectivity in identifying studies, evaluating their relevance, assessing the quality of the methods and extracting results has clear benefits over informal, subjective approaches. It guards against arbitrary selection of studies and allows for replication of at least parts of the review protocol (eg, literature identification). Clearly, meta-analysis has great appeal in the research community. A cursory examination of PubMed searching on ‘epidemiology’ and ‘meta-analysis’ yielded the expected pattern of proliferation—fewer than 100 publications per year prior to 1990, around 400 per year in 2000, 800 per year in 2005, 2000 per year in 2010, 5600 in 2015 and over 6000 per year starting in 2019.

In the 1990s, there was intense debate over the merits and demerits of meta-analysis in observational epidemiology, with some arguing for abandoning this approach entirely<sup>1,2</sup> and others expressing reservations based largely on the heterogeneity of study methods.<sup>3–6</sup> The role of meta-analysis in causal inference specifically also has been considered.<sup>7,8</sup> Interestingly, the debate appeared to end over 20 years ago without a clear resolution, yet meta-analysis became the default approach to summarising and evaluating evidence. We suggest that the debate should be reopened and make the case that the negative features of this approach often outweigh its benefits.

The primary competitor to generating pooled estimates through meta-analysis is

some variant of expert review. Obviously, pooled estimates through meta-analysis and expert reviews may be divergent, deviating in either direction—seemingly clear evidence for an effect based on meta-analysis that is not accounting for important limitations or little evidence in support of an effect from meta-analysis that fails to recognise there are a subset of superior studies that lead to a different conclusion. When the inferences diverge, it is possible that the forced objectivity of the meta-analysis reveals truth that contrasts with the subjective, biased assessment of experts. But the simplification and compromises required to generate pooled estimates may fail to capture important underlying methodological issues that are apparent to expert evaluators. Apparent consistency in results may reflect consistent biases, or there may be a small subset of highly informative studies that are overwhelmed by a large number of weaker ones in a pooled estimate. Examining the impact of study methods on results calls for deep expertise in the subject being evaluated, and meta-analysis is not a reliable substitute for evidence review and synthesis by experts.

## PROBLEMS WITH STRATEGY

### Presumes studies are all approximating the same measure of association

An assumption in generating a pooled estimate across studies is that they should converge on a common causal effect that differs solely from methodological differences across studies. The premise may well be incorrect—the causal impact of a given exposure in a given population may truly be different for a variety of reasons, notably varying prevalence of other causal components.<sup>9</sup>

Even studies of the exact same exposure rarely are addressing the same magnitude and range of exposure when they generate measures of association. Within-study comparisons of higher versus lower exposures often address markedly different absolute levels, and even when we try to scale the association using common metrics, for example, OR per unit of exposure, we may be basing that estimate

on a different region of the dose–response curve where the effect per unit change and extent of measurement error differ.

Finally, the differences in ratio and absolute (difference) measures of effect will influence comparisons across population when the baseline risk of disease differs. If baseline risk differs, it is impossible for both ratio and difference measures of association to be the same across studies: if the difference measures are identical across two such populations, the ratio measures must be divergent, and vice versa.

### Presumes that variation across studies is solely due to random error

Having identified relevant studies, the conventional approach is to generate a precision-weighted estimate of the pooled association. Weighting by precision would be justified if the variation in results were solely a function of random error, yet substantive differences among critical study methods, such as exposure assessment and control for confounding, are much more plausible as explanations of varying results.<sup>6</sup> The heterogeneity in methods provides the richest source of information, more so than a statistically precise estimate of a common effect that ignores that heterogeneity in methods. Over 25 years ago, Greenland<sup>3</sup> clearly stated the problem:

... metaanalysis can only contribute to knowledge beyond single studies if one treats it as a comparative activity, aimed at testing criticisms of study results and identifying patterns or trends in study results. Unfortunately, meta-analysis and pooled-data analyses are most often treated as synthetic exercises, aimed at producing a single summary effect estimate.

## PROBLEMS IN IMPLEMENTATION

### Excludes studies that do not provide the key results needed for pooled estimates

In conducting meta-analysis, only certain types of results can be included, namely those that are sufficiently compatible to be pooled for an aggregate estimate. This can result in the exclusion of studies based on how variables are measured, completeness of documentation in the publications resulting from the study and the methods used to quantify the results. Studies that employ divergent methods can be particularly informative and should not be set aside for convenience. There can also be differences among similar studies due to the data analysis methods, for example, scaling of exposure (arithmetic, logarithmic, dichotomous, multi-level) or the

<sup>1</sup>Department of Epidemiology, Brown University School of Public Health, Providence, Rhode Island, USA

<sup>2</sup>Iriub - National Research Council, Roma, Italy

Correspondence to Professor David A Savitz, Epidemiology, Brown University, Providence, RI 2903, USA; david\_savitz@brown.edu

statistical tool (regression coefficient, ratio or difference measure of effect). Sometimes, the varying statistical tools can be converted to a common scale based on published data,<sup>10</sup> but often that cannot be done and non-compliant studies are simply excluded.

### Heterogeneity is assessed based on study results rather than study methods

Another universally applied tool in meta-analysis is the assessment of statistical heterogeneity of results. When there is compelling statistical evidence of divergent findings among the measures of association across studies, consideration is given to the reasons for that pattern. Without taking the study methods into account, there is little to be learnt from the spread of results and whether it exceeds what would be expected to occur randomly. The important question is whether different study methods generate meaningfully different measures of association, and with that question, useful insights are gained on whether results do or do not diverge in relation to the methods. Different methods that generate similar findings may help to provide assurance that the alternative strategies are not highly influential, and analogously, if different methods generate divergent findings, it suggests that those methods are influential.

### PROBLEMS IN INTERPRETATION Meta-analyses generate the illusion of certainty

Perhaps the single greatest problem with generating pooled estimates of association is the way in which they are interpreted, particularly by audiences that are least familiar with the underlying substantive and methodological issues. While there are often caveats regarding publication bias, heterogeneity or study quality, the fundamental limitations in the approach are often overlooked because of the apparent virtues of including many studies and often remarkable precision of the pooled estimate. Without deeper scrutiny, 'many studies' comes across as a series of replications that collectively reveal the correct value because the studies presumably compensate for one another's shortcoming. The very narrow CIs are

readily construed as suggesting that we have zeroed in on the 'right answer' with certainty. It seems that the audience that is least sophisticated regarding epidemiological methods is most readily persuaded by such results, with more knowledgeable consumers being more sceptical.

### Recommendations

There are a series of relatively simple recommendations that can be offered to refine and narrow the application of meta-analysis for observational studies to serve the interests of improved informed evaluation of causality:

1. Avoid treating meta-analysis as the default approach to assessing a body of research, recognising that it is useful only when a set of similarly designed, but small studies generate imprecise results such that a pooled estimate will provide a clearer indication of the magnitude of association by reducing random error. If random error is not a major source of uncertainty as is the case with large studies, pooling them would only obscure informative heterogeneity based on varying methods.
2. To assess the impact of varying study methods, organise the literature to address specific hypotheses and examine pooled estimates for thoughtfully selected subsets of the literature.<sup>3 11</sup> Concentrate on features that are most likely to result in variation in measures of association specific to the topic.
3. To inform causal inference, consider the magnitude of association in the most valid subset of studies along with evidence pertaining to the presence of substantial bias. In the spirit of the Hill considerations,<sup>12</sup> the measure of association is only the starting point for evaluation.
4. Avoid exclusion of evidence because it is inconvenient to incorporate the data into a pooled estimate, recognising that some informative studies may well not fit into the statistical summary of choice but nonetheless have value and merit inclusion in the overall assessment.

Forgoing some of the elements of neutrality need not result in a completely subjective assessment. A carefully reasoned, explicit (and thus transparent) explanation of why the topic was

examined in the manner decided upon by the experts would yield gains in informativeness that more than outweigh some loss of pure replicability.

**Contributors** The co-authors jointly wrote and edited the commentary and approved the final document.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.



**To cite** Savitz DA, Forastiere F. *Occup Environ Med* 2021;**78**:621–622.

Received 1 May 2021

Accepted 2 June 2021

Published Online First 22 June 2021

*Occup Environ Med* 2021;**78**:621–622.

doi:10.1136/oemed-2021-107702

### ORCID iD

David A Savitz <http://orcid.org/0000-0002-3475-4113>

### REFERENCES

- 1 Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;140:771–8.
- 2 Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997;50:223–9.
- 3 Greenland S. Can meta-analysis be salvaged? *Am J Epidemiol* 1994;140:783–7.
- 4 Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;140:290–6.
- 5 Feinstein AR. Meta-Analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;48:71–9.
- 6 Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. *Am J Epidemiol* 1995;142:371–82.
- 7 Weed DL. Meta-Analysis under the microscope. *J Natl Cancer Inst* 1997;89:904–5.
- 8 Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol* 2000;29:387–90.
- 9 Pearce N. Epidemiology in a changing world: variation, causation and ubiquitous risk factors. *Int J Epidemiol* 2011;40:503–12.
- 10 Steenland K, Barry V, Savitz D. Serum perfluorooctanoic acid and birthweight: an updated meta-analysis with bias analysis. *Epidemiology* 2018;29:765–76.
- 11 Savitz DA, Wellenius GA, Trikalinos TA. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. *Am J Epidemiol* 2019;188:1581–5.
- 12 Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med* 2015;108:32–7.