

What's with epidemiologic meta-analyses?

Overemphasis on individual studies and false alarms are pushing epidemiology to its limits, claims a *Science* news report (1). The solution is, many think, aggregation of results from studies that address the same hypothesis. Power considerations, the argument goes, encourage the consolidation of study results so that moderate effects can be detected. But if results of individual studies differ widely, should one aggregate?

Further down the line, others marvel whether meta-analyses should be conducted in nonexperimental research at all. I do not see much of a problem in well-conducted nonexperimental meta-analyses, to the extent the data at hand are combinable, do not suffer from excessive publication bias, and allow for the reduction or prediction of heterogeneity.

Next to publication bias, heterogeneity is a major reason for skepticism towards nonexperimental (perhaps also experimental) meta-analyses. Heterogeneity refers to variation in estimates from different studies of a single parameter value of the average effect or exposure-response gradient (fixed effect) or the variation in effect parameters themselves (random effects, "true" heterogeneity) or a combination of the two. There is a host of sources for heterogeneity: conceptual targeting and operationalizations of determinants and end points, study type, modification, confounding, reference entities for the exposed or for cases, publication bias, chance, and others. Different indicators of exposure are applied, implying differences in exposure circumstances, levels and durations, cumulated exposures, exposure categorizations and cut points, and concomitant confounding exposures. Variations in determinant levels of doses may offer a "natural" explanation for heterogeneity (2) and warrant an exposure-response analysis.

There is an additional level of heterogeneity that is less formal in definition but that deserves equally serious attention. This is the occasional case of differing results of meta-analyses that address the same phenomenon or hypothesis.

Why different results from meta-analyses? Apart from the trivial but important case of analyses upgraded with new data, the question boils down to the selection of studies and, within studies, to the selection of inputs into meta-analyses. Are these matters of conscious or unconscious patterns of preference? Definitely yes, what else. It is believed that subjectivity is relieved by reaching a consensus between several evaluators. Subjectivity, however, is not restricted to an individual assessor. It may well creep into a group of evaluators, even when they follow well-defined rules of extraction. One meta-analyst or group discards a somewhat low-quality, off-target, or unpublished study, while another includes it. A sloppy search, even database-oriented, may identify only half of the available studies. When it comes to individual studies, one is usually faced with an array of rate ratios in a single study. One might pick those that carry maximal statistical weight, while another would prefer unconfounded and lagged estimates and write to the authors to improve targeting. One might lump categories of high-level and long-duration exposures together into a "substantial" category, while another might be careful not to do so.

Carefully conducted meta-analyses of randomized clinical trials are worth studying by epidemiologists who consider data aggregation. A recent update on chemotherapy in nonsmall-cell lung cancer (3) serves as an example, even though clinicians might consider the estimated benefit of chemotherapy rather on the meager side. One might also learn about possibilities that modeling offers for the reduction of heterogeneity and the adjustment for components of study quality.

It is obvious that meta-analysis is not an automatic answer to questions pertaining to multiple data, for study results may not be combinable. In nonexperimental epidemiology, valid exposure reconstruc-

tion and meaningful exposure conceptualization and categorization should, by today, have become a sine qua non in occupational, environmental, nutritional, and life-style studies, easing the task of the meta-analyst. Still, exposure categories may cover forbiddingly wide ranges of exposure, often because nonexperimental exposure circumstances are just that way. Categories may overlap on account of misclassification. It may be worthwhile discarding studies with fuzzy exposure indicators, or see how sensitive the meta result is to these studies.

In spite of variations in the ways meta-analyses are carried out, I am sometimes struck by a scenario in which one day someone from outside academic epidemiology stands up declaring that epidemiologic meta-analyses should be canonized into a set of rules, which are then listed in a legal order and represent supposedly Universal Good Meta-Analytic Practice to be followed henceforth. I just do not believe in a one-and-only normative meta-analytic algorithm, or in any eternal epidemiologic algorithm, for that matter.

Finally, there are three good alternatives to meta-analysis: not to aggregate, analysis of pooled data, and multicentric studies. The null option is fine in situations in which there is nothing to aggregate or when it turns out that the studies that were meant to be aggregated are biased or not combinable, or when one simply does not believe in aggregating any data.

Pooled studies go back into the primary data of a number of studies, into individual records of subjects. Pooling may make it possible to employ sharper common exposure dimensions and categories, compared with meta-analysis. Modern modeling techniques may be applied to the pooled data set. Control of confounding and scrutiny of modification may become much more powerful than in conventional meta-analyses. A recently completed study of 28 704 wood and furniture workers (4) represents an instructive reanalysis of five studies dealing with cancer mortality among workers in wood-related industries.

Finally, multicentric studies offer the best possibilities for meaningful data aggregation. The key element is a common protocol, based on a feasibility study, the latter often accompanied by a pilot phase. The protocol is prepared and agreed upon before the data collection. The International Agency for Research on Cancer conducts such powerful multicentric studies. These studies are geared towards priority issues such as the carcinogenicity of occupationally occurring exposures to styrene or bitumen fumes.

A final consideration between no aggregation, meta-analysis, pooled reanalysis, and multicentric studies is cost. I wonder if I go totally astray by advancing a guess that the respective costs of these alternatives will be roughly proportional to 0:1:10:100 on the average. Traditional verbal review would probably weigh in at about 0.5. One might also consider a smart combination of an in-depth narrative review, strengthened with a formal meta-analysis, at a cost not exceeding 1.5 on this scale.

References

1. Taubes H. Epidemiology faces its limits. *Science* 1995;269:164—9.
2. Hertz-Picciotto I, Neutra RR. Resolving discrepancies among studies: the influence of dose on effect size. *Epidemiology* 1994;5:156—63.
3. Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *Br Med J* 1995;311:899—909.
4. Demers P, Boffetta P, Kogevinas M, Blair A, Miller BA, Robinson CF, et al. Pooled reanalysis of cancer mortality among five cohorts of workers in wood-related industries. *Scand J Work Environ Health* 1995;21:179—90.

Timo Partanen
Department of Epidemiology and Biostatistics
Finnish Institute of Occupational Health
Topeliuksenkatu 41 a A
FIN-00250 Helsinki Finland