VALUE IN HEALTH 14 (2011) 414–416

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jval

EDITORIAL

ISPOR States Its Position on Network Meta-Analysis

Indirect comparisons and network meta-analysis are being seen increasingly often in cost-effectiveness analyses, reimbursement decisions, and academic journals. In essence, they allow investigators to draw coherent conclusions about the comparative efficacy of any number of treatments, based on evidence from randomized trials, which normally compare only two or three treatments. Statistical methods for network meta-analysis [1] go back to Gleser and Olkin [2], Higgins and Whitehead [3], and Hasselblad [4]. The idea had appeared earlier in the Confidence Profile Method literature [5], but not in a form compatible with the accepted principles of meta-analysis. The two-part report [6,7] from ISPOR's Task Force on Indirect Comparisons seems to represent the first position statement from an academic body on these methods. Prepared by a group drawn from major consultancy companies, the pharmaceutical industry, and academia, they provide a strong but balanced endorsement of the methods, particularly the Bayesian forms of analysis that fit conveniently within the probabilistic decision-modeling framework. At the same time, the report presses for more research to extend the methods into further areas of secondary analysis, such as synthesis of multiple correlated outcomes and to covariate adjustment, or meta-regression, while recognizing the difficulties of reliable covariate adjustment in the sparse data sets usually available.

ISPOR's position contrasts with the more cautious approach evident in the methodology guidelines published by some of the reimbursement authorities. The UK's National Institute for Health and Clinical Excellence (NICE) [8] made "direct evidence" the base case in appraisals of new technologies but allowed combination of "direct" and "indirect" as a secondary analysis. In Australia, the Pharmaceutical Benefits Advisory Committee makes limited use of network synthesis because its procedures encourage identification of a single comparator [9]. The Canadian Agency for Drugs and Technologies in Health [10] also adopts a cautious stance. The 2008 Cochrane Handbook [11] avers that "indirect comparisons may suffer the biases of observational studies" and advises that direct and indirect evidence should only be combined as a supplemental analysis.

One might wonder: what was the basis for the caution previously expressed by leading experts? Or has the ISPOR Task Force discovered something previously overlooked? In truth, it is probably the sheer inevitability of network meta-analysis that is the decisive factor. If only two interventions are involved, one might debate whether indirect evidence—the use of A versus B and B versus C trials to infer an A versus C treatment effect—should be used. But as soon as a decision has to be made between more than two treatments, estimates of their comparative efficacy must have the property of coherence. This can be represented as follows. If $\hat{d}_{XY}$ is an estimate of the effect of Y relative to X on an appropriate scale, then coherent estimates must conform to:

$$\hat{d}_{AC} = \hat{d}_{AB} + \hat{d}_{BC} \quad (1)$$

Indirect comparisons and network meta-analysis are, in the end, simply methods that deliver relative effect estimates that are "coherent" in this sense. Without this property, coherent decision making would be impossible. If Peter is 3 inches taller than Paul, and Paul is 2 inches taller than Mary, we cannot allow that Peter can be 6 inches taller than Mary. If one believed this could be possible, there could not be a reliable way of deciding who was tallest.

Thus, whatever the misgivings, the inexorable logic of coherent comparison must lead us back to Equation (1). It comes as no surprise, therefore, that whatever may be written in the NICE Methods Guide, which are anyway a little ambiguous [12], coherent estimates of effects achieved via network meta-analysis have become a familiar feature of technology appraisals at NICE, just as they make a frequent appearance in efficacy comparisons in leading clinical journals.

Can it all be that simple? It would be, but for heterogeneity.

As an area of statistical enquiry, the development of meta-analytical method and practice has been rather strange. The statistical literature does not lack for theoretical analyses: methods for combining 2 × 2 tables go back more than 50 years. But there has been a gap between what we may call the "scientific" exercise, which sets out to answer questions like "which is the best treatment for patient group X?," and the "literature summary" exercise, which asks "what does the literature say about treatments A, B, and C?" Distinguished voices have voted for science over summary [13], with the random-effects model attracting some particularly critical comments [14]. On the whole, their advice has been ignored, and the great majority of systematic reviews and meta-analyses are summary, not science. The key driver for decision makers and clinicians is that there must be a clear definition of the target patient group to get an answer to the scientific question. This remains true regardless of how much evidence there is specifically on target population X or whether—as can often be the case—most of the evidence turns out to be on similar group Y. The literature summarizer could, of course, have the same definition as the scientist, but never *needs* to have one.

Health technology assessment, of course, stands on the science side. With network methods, the question of how wide to draw the target population can translate into how wide to make the comparison network. The report refers to work by Hawkins et al. [15], outlining a structured approach to defining the size of the comparison network, but a key section correctly sees the trade-off between increased precision and robustness that follows from the larger evidence base against the increasing danger of *heterogeneity*, as an ever wider class of patient groups is thrown into the analysis. Throughout the report, the danger to network meta-analysis arising from the presence of effect mod-

ifiers—the cause of heterogeneity—is repeatedly emphasized. If there is an imbalance in the presence of effect modifiers in AB comparisons and AC comparisons, then any conclusions about the BC contrast could be in jeopardy.

However, heterogeneity continues to be the source of confusion in both the meta-analysis community and among health economists, and the ISPOR report is no exception. The authors propose the curious idea that although heterogeneity could introduce confounding into network meta-analysis, which is true, it may actually be an advantage in pairwise comparisons, where it might "increase generalizability" and "may be welcome if it reflects real world practice." This is an unusual view: most authorities have argued at length that heterogeneity needs to be accounted for and removed [11,16], for example, by adjusting for effect modifiers and for sources of random bias [17]. Clearly, enlarging the number of *trials* increases robustness, but heterogeneity only increases the *difficulty* of generalizing from the evidence as it is given to the target population. All meta-analytic estimators are weighted averages of the observed trial-specific treatment effects, and if the latter are viewed as unbiased, then the weighted average will be unbiased as well. The weighted average of several heterogeneous, but unbiased, estimates is also unbiased, but it becomes difficult to say what it is an unbiased estimate *of*. As the between-study variation increases, the relevance of the evidence to the specific target population for decision becomes less and less certain—regardless of our certainty about the *mean* effect.

Like so much else in the area of meta-analysis, much of what passes for good practice, including most famously the "hierarchy of evidence" [18], is based on rather rough-and-ready common sense. Sometimes a more formal analysis would be helpful. The ISPOR report repeats the view, passed around rather uncritically, that network meta-analysis makes the assumptions of "homogeneity," "similarity," and "consistency." Fortunately, there is some formal work to which we can appeal. First, suppose that in a pairwise meta-analysis of n_{AB} trials comparing A and B, the true treatment effect $\delta_{j,AB}$ in trial j is sampled from a common distribution $\delta_{j,AB} \sim N(d_{AB}, \sigma^2)$, and in a pairwise meta-analysis of n_{AC} A *vs*. C effects, the true effects are from another distribution $\delta_{j,AC} \sim N(d_{AC}, \sigma^2)$. The *only* additional assumption in a network synthesis is that each of these exchangeability assumptions extends over the entire set of $(n_{AB} + n_{AC})$ trials. If BC trials are now conducted and the same *two* assumptions are now extended to the additional n_{BC} trials, then the requirement for "consistency" of the AB, AC, and BC effects, which confers the required coherence property of network estimates (1), is *automatically* met [19] without further assumptions. Thus, the assumptions of network meta-analysis, are, with just that one small extension, no different from the assumptions that have been made all along in pairwise meta-analysis. Furthermore, it can be shown that the coherent estimates produced by network meta-analysis are, like all meta-analytic estimates, weighted averages of the study-specific estimates from the original trials [20].

These theoretical results confirm that network meta-analysis makes no extra assumptions; it simply extends the standard ones. Of course, this in no way detracts from the warnings that many experts have sounded about inference from indirect comparisons and network meta-analysis, regarding the need to assess the validity of assumptions statistically [21] and—perhaps more importantly—clinically. Indeed, it means that the new network methods inherit all the *problems* of pairwise meta-analysis and suggests that the caveats routinely expressed about network meta-analysis should have been applied all along to pairwise comparisons in which, as noted previously, the culture of "literature summary" has encouraged an excessive use of random effects to average over clinically significant heterogeneity.

How should the practice of economic evaluation respond to unexplained variation in treatment effects? A suggestion originally made by Spiegelhalter et al. [22] and subsequently taken up by others [17,23] has been to avoid using the random-effects mean in economic evaluation. Except under some very specific interpretations, it conveys a far greater degree of certainty than can be realistic if different trials estimate widely different effect sizes. Instead, it is proposed that modelers should use the predictive distribution of the effect in a future study. This increases the uncertainty around the estimated effect very considerably, in effect adding the between-trials variance to the variance of the mean effect. Probably this approach would confer an excessive level of uncertainty on treatment efficacy estimates because it assumes that *all* effect variation is due to true variation in treatment effects rather than a mixture of true variation and random biases internal to each trial [17,24]. Either way, the response to unexplained heterogeneity is an issue that requires urgent consideration within the synthesis community and among health economists. The probability that a treatment is the most effective, the probability that it is cost-effective, cost-effectiveness acceptability curves, and the expected value of further research [25] will all be highly sensitive to the outcome of this overdue debate.

A. E. Ades, PhD
School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

1098-3015/$36.00 – see front matter

Published by Elsevier Inc.
doi:10.1016/j.jval.2011.05.001

REFERENCES

[1] Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21:2313–24.
[2] Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds., The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994.
[3] Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;5:2733–49.
[4] Hasselblad V. Meta-analysis of multi-treatment studies. Med Decis Making 1998;18:37–43.
[5] Eddy DM, Hasselblad V, Shachter R. Meta-analysis by the Confidence Profile Method. London: Academic Press, 1992.
[6] Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on indirect treatment comparisons good research practices: part 1. Value Health 2011;14: 417–28.
[7] Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. Value Health 2011;14:429–37.
[8] National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. London: NICE, 2008.
[9] Pharmaceutical Benefits Advisory Committee. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee, Version 4.3. Canberra: Australian Government, Department of Health and Aging, 2008.
[10] Wells GA, Sultan SA, Chen L, et al. Indirect Evidence: Indirect Treatment Comparisons in Meta-analysis. Canadian Agency for Drugs and Technologies in Health, 2009.
[11] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: The Cochrane Collaboration, Wiley, 2008.
[12] Sculpher MJ. NICE's 2008 Methods Guide: sensible consolidation or opportunities missed. Pharmacoeconomics 2008;26:721–4.
[13] Rubin DB. A new perspective on meta-analysis. In: Wachter KW, Straf ML, eds., The Future of Meta-analysis. New York: Russell Sage Foundation, 1990.

[14] Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol 1994;140:290–6.
[15] Hawkins N, Scott DA, Woods B. How far do you go? Efficient searching for indirect evidence. Med Decis Making 2009;29:273–81.
[16] Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. In: Chalmers I, Altman DG, eds., Systematic Reviews. London: BMJ Publishing Group, 1995.
[17] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Series (A) 2009;172:137–59.
[18] McAlister F, Laupacis A, Wells GA, Sackett DL. Users' guides to the medical literature. XIX Applying clinical trial results. B. Guidelines for determining whether a drug is exerting (more than) a class effect. JAMA 1999;282:1371–7.
[19] Lu G, Ades AE. Modelling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009;10:792–805.
[20] Lu G, Welton NJ, Higgins JPT, et al. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. Res Synthesis Methods (in press).
[21] Song F, Loke Y-K, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009;338:b1147.
[22] Spiegelhalter DJ, Abrams KR, Myles J. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. New York: Wiley, 2004.
[23] Ades AE, Lu G, Higgins JPT. The interpretation of random effects meta-analysis in decision models. Med Decis Making 2005;25:646–54.
[24] Welton NJ, Ades AE, Carlin JB, et al. Models for potentially biased evidence in meta-analysis using empirically based priors. J R Stat Soc Series (A) 2009;172:119–36.
[25] Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. Health Econ 1996;5:513–24.