VALUE IN HEALTH 18 (2015) 673–681

Available online at www.sciencedirect.com

ScienceDirect

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

Extending Treatment Networks in Health Technology Assessment: How Far Should We Go?

Deborah M. Caldwell, PhD, Sofia Dias, PhD, Nicky J. Welton, PhD*

School of Social and Community Medicine, University of Bristol, Bristol, UK

ABSTRACT

Background: Network meta-analysis may require substantially more resources than does a standard systematic review. One frequently asked question is "how far should I extend the network and which treatments should I include?" **Objective:** To explore the increase in precision from including additional evidence. **Methods:** We assessed the benefit of extending treatment networks in terms of precision of effect estimates and examined how this depends on network structure and relative strength of additional evidence. We introduced a "star"-shaped network. Network complexity is increased by adding more evidence connecting treatments under five evidence scenarios. We also examined the impact of heterogeneity and absence of evidence facilitating a "first-order" indirect comparison. **Results:** In all scenarios, extending the network increased the precision of the A versus B treatment effect. Under a fixed-effect model, the increase in precision was modest when the existing direct A versus B evidence was already strong and was substantial when the direct evidence was weak. Under a random-effects model, the gain in precision was lower when heterogeneity was high. When evidence is available for all "first-order" indirect comparisons, including second-order evidence has limited benefit for the precision of the A versus B estimate. This is interpreted as a "ceiling effect." **Conclusions:** Including additional evidence increases the precision of a "focal" treatment comparison of interest. Once the comparison of interest is connected to all others via "first-order" indirect evidence, there is no additional benefit in including higher order comparisons. This conclusion is generalizable to any number of treatment comparisons, which would then all be considered "focal." The increase in precision is modest when direct evidence is already strong, or there is a high degree of heterogeneity.

Keywords: comparative effectiveness, health technology assessment, literature searching, mixed treatment comparisons, network meta-analysis, systematic review.

Introduction

Indirect comparisons and network meta-analysis (NMA) are increasingly common in the evaluation of multiple competing health technologies when interest lies in the relative rankings of all treatments of clinical interest [1]. NMA is also used by health reimbursement agencies worldwide, including the National Institute for Health and Care Excellence (NICE) Single Technology Appraisals (STAs) program, where the objective is to assess whether a treatment should be available for use on the National Health Service in England and Wales. STAs are the mainstay of the NICE health technology assessment (HTA) program; of the 33 appraisals published in 2013, 29 were completed under the STA process (www.nice.org.uk). STAs typically evaluate a single treatment close to marketing launch, and as such the focal comparison of interest is with standard/usual care options. We note that this is true even when multiple treatments are included in a network and relative rankings reported [2].

NMA may be used in STAs when direct evidence from trials of A versus B is either unavailable or sparse; however, no formal guidelines exist to ensure transparency on which treatments should be included, when to extend a network, or how far it should be extended. In the absence of such guidelines, there are concerns that networks could be defined specifically to favor a particular treatment [3,4]. Proposals for the assessment of network geometry have received attention [5,6], and network size has been described as an "unsolved issue" in NMA [7]. In an empirical study of 18 published networks, Mills et al. [8] examined the impact of retrospectively excluding treatments and note how treatment effect estimates and treatment rankings were modified. In STAs, however, the starting network consists of a *fixed* "decision set" of treatments (i.e., treatment and comparator (s) of interest) to which additional evidence (a "supplementary set" of treatments identified a priori) may be prospectively included to connect those already in the network. Such an approach has been separately described by Ades et al. [9] and

* *Address correspondence to:* Deborah M. Caldwell, School of Social and Community Medicine, University of Bristol, 39 Whatley Road, Bristol BS8 2PS, UK.

E-mail: d.m.caldwell@bristol.ac.uk.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2015.03.1792

Hawkins et al. [10] and is referenced by ISPOR Task Force [11] and NICE methodology guidelines [12].

A recent case study calls for further work to evaluate network size and structure and provide generalizable findings on the added value of extending treatment networks [13]. Indeed, there is a practical need to ask how far to extend a network in STAs [14], what is the benefit of doing so, and whether there is a diminishing return for including additional treatments. NMA is understood to be more resource intensive than traditional pairwise systematic review [15]. For example, literature searching, screening, eligibility assessment, and data extraction may be more cumbersome because of the increased number of studies to review, although this will vary depending on the network. The further a network is extended, the risk of bias, heterogeneity, and inconsistency may also increase. This would further add to the reviewer's workload assessing whether the assumption of consistency/transitivity holds across the network [16]. However, previous empirical work suggests that combining direct and indirect evidence may increase the precision of treatment effect estimates across a network [17]. Taking the perspective that the purpose of evidence synthesis is to reduce uncertainty in decision making, a key consideration in the development of guidelines on how far to extend evidence networks is the impact on the precision of the focal treatment comparison(s).

In this article, we explore the effect of combining direct and indirect evidence in an NMA on the precision of a single pairwise comparison in a hypothetical six-treatment network. Our starting point is to assume that a literature search has been conducted and has generated a "star"-shaped starting network. We explore the effects of "extending" the network by including additional evidence situated at different points in the network. The article is structured as follows. First, we define the statistical properties of indirect comparisons. Then, we introduce the network structure and describe the different evidence scenarios considered here. The statistical method is described and findings are reported. We conclude by discussing the practical implications of the findings, make recommendations for the systematic review component of HTA, and discuss implications for NMA, in general.

Methods

In a three-treatment network, an indirect estimate of the A versus B treatment effect estimate is derived as follows:

$$\theta^{I}_{AB} = \theta^{D}_{AC} - \theta^{D}_{BC} \quad (1)$$

where θ represents a treatment effect estimate (e.g., log-odds ratio, mean difference) and where superscript I denotes an indirect estimate and superscript D denotes a direct estimate. The variance of θ^{I}_{AB} is equal to the sum of the variances $\hat{V}^{D}_{AC}$ *and* $\hat{V}^{D}_{BC}$ estimated from the direct A versus C and B versus C comparisons, $\hat{V}^{I}_{AB} = \hat{V}^{D}_{AC} + \hat{V}^{D}_{BC}$. Here, we define A and B as our focal treatments of interest. Any comparison of A or B to another treatment (e.g., C) is defined as contributing "first-order" evidence if it facilitates a triangular loop (e.g., A vs. C and B vs. C) [10]. A comparison that does not include either A or B but that facilitates a quadrilateral loop of evidence (e.g., C vs. D in the loop A-B-C-D) is defined as providing "second-order" evidence for the focal treatments of interest A and B.

Network Formation

Our starting point was to assume that a literature search has been conducted and has generated a network with six treatments labeled A, B, …, F, where treatments A and B form the "decision set" of treatments and the effect estimate of interest is θ_{AB}. For simplicity, we assume a known network size, such that all possible comparisons can be known a priori. Six is the median number of treatments observed in published NMAs [18]. In a standard systematic review, only direct evidence on contrast A versus B (Fig. 1A) would be reported, which represents a single pairwise meta-analysis here. Note that the solid lines connecting each pair of treatments in Figure 1 indicate that there is direct evidence available for that contrast. Drawing on the principles of an iterative strategy for NMA [10], we assume that evidence "closest" to the focal treatment comparison of interest will be included first. Here we first add evidence on all comparisons including treatment A, forming a "star" network structure (Fig. 1B). We then add evidence that forms triangular "first-order" loops for A versus B (B vs. C, B vs. D, B vs. E, and B vs. F) [19] (Fig. 1C,D). Second-order indirect evidence, via treatment C, is added next (Fig. 1E). The final level of network complexity (Fig. 1F) is to include all evidence via D versus E, D versus F, and E versus F.

Description of Evidence Scenarios

(i) Network with Evidence Available for All Contrasts

Here we concentrate on a network structure in which direct evidence is available for θ_{AB}, albeit in differing amounts. Five *hypothetical* scenarios are considered under an assumption of consistency (Equation 1). In each scenario, we assume that values for the observed precision of treatment effect estimates are available for every pairwise contrast. The resulting precision of the pooled NMA estimate for A versus B depends only on these input precisions and not on the actual observed treatment effects (see Appendix 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.03.1792). No assumptions are made about the observed treatment effects, and results are general for any outcome measure with our assumed input precisions. Furthermore, our conclusions are based on the relative precision across different parts of the network, rather than on the absolute value. Input precision values for each scenario are reported in Table 1.

Scenario 1: *Equal variance is assumed for each contrast across the network*. Here, each contrast θ_{XY} is informed by a meta-analysis with variance, $V_{XY} = 1$, where V_{XY} is the observed variance (SE^2) from a meta-analysis of X versus Y. The precision of X versus Y is defined as $P^{D}_{XY} = 1/V_{XY}$.

Scenario 2: A versus B comparison is the "weakest" link in the six-treatment network. Contrasts contributing first-order indirect evidence are also weak (imprecise), and second-order contrasts contribute even weaker evidence for A versus B. This scenario is sometimes seen when fewer trials are conducted for ethical or practical reasons, for example, in pain management for women in labor [20]. Note that values assigned in all scenarios are hypothetical, and do not exactly replicate the illustrative HTAs.

Scenario 3: The A versus B comparison is the "weakest" link in the six-treatment network, with the contrasts forming both first- and second-order indirect comparisons being stronger. In HTA, this scenario is seen when A versus B are interventions from rival manufacturers that have seldom been compared, or are compared only in a small study [21]. Evidence in such networks is likely to be found on the newer technologies versus placebo/standard care and on the standard versus older interventions.

Scenario 4: A versus B is the strongest link in the six-treatment network, with the contrasts forming indirect comparisons being weaker. This scenario may be seen in practice when both A and B are older interventions, perhaps the criterion standards for the clinical area, and have been trialed many times [22].

Scenario 5: A versus B comparison is the strongest link, with the contrasts contributing to indirect comparisons also being strong. This scenario may be seen in practice with "me-too"