

EDITORIAL COMMENT



Do We Know the Best Treatment for In-Stent Restenosis Via Network Meta-Analysis (NMA)?

Simple Methods Any Interventionalist Can Use to Assess NMA Quality and a Call for Better NMA Presentation*

Stephen G. Ellis, MD

In this issue of *JACC: Cardiovascular Interventions*, Lee et al. (1) present the results of a Bayesian network meta-analysis (NMA) to assess the relative efficacy of drug-eluting stents (DES), drug-eluting balloons (DEB), and plain old balloon angioplasty (POBA) used to treat in-stent restenosis (ISR), using target lesion revascularization (TLR) as the outcome measure and 11 randomized, controlled trials as the data source. They conclude that both DES and DEB reduce TLR by ~75% to 80% compared with POBA. This is supported by data from angiographic follow-up. Safety issues (death and myocardial infarction) were also considered. They go on to assess the probability of the best treatment and conclude that it is most likely to be treatment with DEB (60% likelihood).

SEE PAGE 382

I suspect most cardiologists experience some form of “brain freeze” when confronted with the results of an NMA, and I suspect most of us simply jump to the conclusions and hope the paper was well enough refereed so they are valid. Sadly, that may not always be true (2).

Fortunately, there are several excellent and readable reviews on the topic, not only in statistical journals (3-6), but also mainstream journals such as

the *British Medical Journal* (7) and *Annals of Internal Medicine* (8,9) that demystify the topic.

In short, NMAs are used to draw comparisons between outcomes of treatments that have been either directly or indirectly compared by making use of comparison treatments they share in common.

Here is a simple step-by-step way to begin to judge the strength of evidence and some suggestions about how to present NMAs to make it even easier to make that assessment (skip to the last paragraph if you have no interest in NMAs).

1. Almost all NMA papers contain a graph showing “network geometry”—the treatments analyzed and connectivity of the trial data used (see Figure 1 in the Lee et al. (1) paper). Check the number of lines (trials) between the nodes (treatments to be compared) and whether the lines directly connect the trials or run through other trials. The more the connecting lines and more direct the connections, the better the strength of evidence is.
2. Downgrade comparisons made primarily by underpowered trials (graphic presentation can be improved to make this easy to visualize; see Figure 1).
3. Heterogeneity of trial design and results are killers for NMA. Differences in inclusion criteria or treatment application (e.g., low risk vs. all-comers DES trials, heparin dose, and bivalirudin treatment duration in ST-segment elevation myocardial infarction medication trials) can make the results of indirect comparisons very different from those of direct comparisons (which obviously do not suffer have this problem). This is termed *incoherence*. Check to see that authors

*Editorials published in *JACC: Cardiovascular Interventions* reflect the views of the authors and do not necessarily represent the views of *JACC: Cardiovascular Interventions* or the American College of Cardiology.

From the Cleveland Clinic, Cleveland, Ohio. Dr. Ellis has reported that he has no relationships relevant to the contents of this paper to disclose.

specifically deal with this issue by comparing results of direct and indirect comparisons between 2 treatments (when possible). Heterogeneity of trial results per se should also be assessed by Cochran's Q or I^2 testing (significant differences are suggested by values of $p < 0.1$ for simple NMAs or $p < 0.15$ for complex NMAs, or $I^2 > 50\%$, respectively). Authors can make this easier by modifying their network geometry graphs (Figure 1).

4. Make sure random-effects modeling, rather than fixed-effects modeling was used (the latter assumes that there is a negligible amount of heterogeneity, which is usually incorrect).
5. Be skeptical about claims for ranking treatments in order of effect because as they are particularly prone to error (7).

Suggestions for better "network geometry" graphic depiction (Figure 1) are the following:

1. Distinguish trials adequately powered for the outcome of interest (e.g., dark lines) from underpowered trials (e.g., lighter lines) (some authors do this already).
2. Distinguish trials with the exact population of interest and same design (e.g., solid lines) from trials

with slightly different populations or design (e.g., dotted lines). See the issue of incoherence discussed previously.

A suggestion for a better "effects summary" depiction is to note to the side of each estimate of effect the strength of evidence on which the estimate rests (strong, medium, weak).

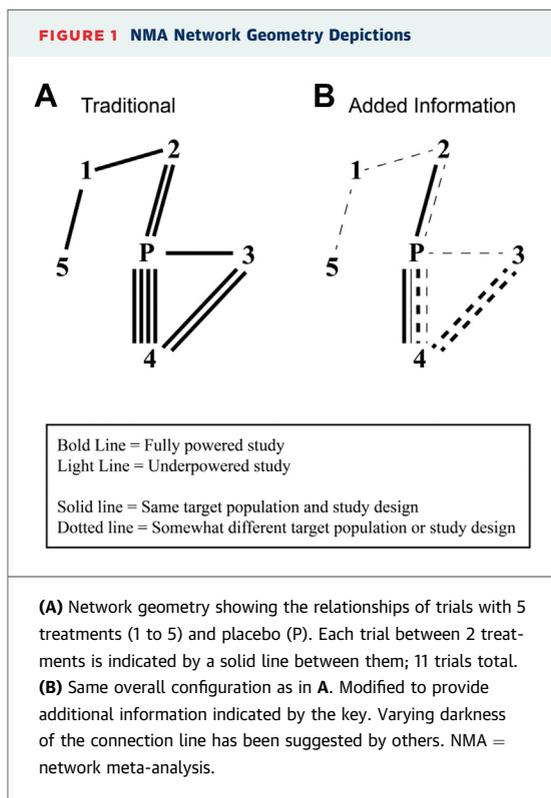
With that said, clearly there are other nuisances of NMAs to consider, including more detailed assessment of the quality of the trials included, and a good statistician should be asked to referee papers with NMAs.

Back to the issue at hand, what does this analysis add to our understanding of how to treat ISR? First, the authors' general approach to NMAs (searching PubMed and other websites, using a Bayesian approach with WinBUGS software) is standard in the field. Their use of a random-effects model, comparison of direct- and indirect-effect estimates using node-splitting, reporting statistical heterogeneity with I^2 , and sensitivity testing is laudable.

Second, reviewing their Figure 1B, one sees at least 3 trials bridging each treatment, which is strength.

Conversely, there are some issues of concern. First, including underpowered trials undercuts the reliability of the results. Because of the large expected difference in TLR between POBA and the other treatments, trials comparing these treatments can be small (e.g., 43 patients per group, assuming 40% vs. 10% TLR with 85% power). However, DEB versus DES trials, attempting to see differences on the order of 15% versus 10% with the same power, need to include ~820 patients in each group, more patients per group than are available in aggregate in the NMA that we are evaluating. Statistical power for low-frequency events such as death and myocardial infarction is even worse. Finally, NMAs combining results with some trials having no events at all (e.g., no deaths or no myocardial infarctions) are notorious for unstable p values.

Second, there are multiple reasonable concerns that can be raised about heterogeneity that weaken their results. For *incoherence*, differences in study design including treating both BMS and DES ISR using a variety of different DES (principally first-generation DES) and lumping their results together, mixing TLR and target vessel revascularization, and using data from differing lengths of follow-up all are problems. Further, patient level data from which to adjust for confounders was not available. With regard to *statistical heterogeneity*,



for many comparisons, I^2 values exceed the 50% threshold.

To summarize, the magnitude of DEB and DES benefit for TLR compared with POBA is such that, despite the concerns raised, it is most certainly correct in order and general magnitude. Conclusions regarding the relative efficacy and safety of DEB and

DES in this circumstance, particularly generalizing to current devices, are premature.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Stephen G. Ellis, Cleveland Clinic, 9500 Euclid Avenue, J2-3, Cleveland, Ohio 44195. E-mail: elliss@ccf.org.

REFERENCES

1. Lee JM, Park J, Kang J, et al. Comparison among drug-eluting balloon, drug-eluting stent, and plain balloon angioplasty for the treatment of in-stent restenosis: a network meta-analysis of 11 randomized, controlled trials. *J Am Coll Cardiol Intv* 2015;8:382-94.
2. Sang F, Lake YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009;338:b1147.
3. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Methods* 2002;21:2313-24.
4. Salanti G, Higgins J, Ioannidis J. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2007;17:279-301.
5. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008;26:763-7.
6. Hoaglin DC, Hawkins N, Janssen 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.
7. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
8. Moher D, Liberati A, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Ann Intern Med* 2009;151:264-9.
9. Cipriani A, Higgins PT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159:130-40.

KEY WORDS drug-eluting balloons, drug-eluting stents, in-stent restenosis, network meta-analysis