# Tangled in the Network: Challenges for Indirect Treatment Comparisons in NICE HTA Submissions

*Alison Martin, MB ChB, Senior Research Scientist; Maria Rizzo, MSc, Senior Research Associate; Ike Iheanacho, BSc, MB BS, Research Scientist and European Director, Health Economics and Epidemiology*

Manufacturers' submissions for reimbursement to the National Institute for Health and Care Excellence (NICE) commonly include network meta-analysis (NMA) evidence suggesting how their product's efficacy relates to that of comparators. NMAs include indirect and mixed treatment comparisons, which permit an estimate of how two or more interventions might compare in the absence of robust head-to-head trials. The reimbursement decision may depend heavily on what NICE makes of the rationale, methodology and results of these analyses. So any manufacturer planning to submit NMA-based evidence needs to know how NICE has faulted such data. With this in mind, we present here an overview of these criticisms.

## BACKGROUND

NICE requires manufacturers to submit a review of the efficacy and safety and a cost-effectiveness analysis for their new drug compared with treatments currently used for the target population. Ideally, these assessments should be based on head-to-head randomised controlled trial (RCT) comparisons between the manufacturer's drug and these other interventions. However, such direct evidence is often absent for key comparisons. In these cases, NICE expects manufacturers to conduct a series of pairwise meta-analyses, together with an NMA, if appropriate, to estimate the likely results of head-to-head comparisons. An NMA can include both direct evidence and indirect comparisons for interventions that have not been directly compared in an RCT but are joined in a network through a common comparator.

Crucially, NMAs should be based on RCT data from similar populations and, like other meta-analyses, they are undermined if there are significant systematic differences between the included trials, since this heterogeneity can cause discrepancies in the studies' respective estimates of the effectiveness of interventions. NICE therefore requires manufacturers to identify heterogeneity between RCTs in an NMA and determine consistency between effect sizes seen in these studies and estimates from the indirect evidence in the network, and that any substantial heterogeneity is explored using meta-regression or subgroup analysis.

## NMAs WITHIN NICE SUBMISSIONS

NICE publishes approximately 30 appraisals each year on new technologies seeking market access in England and Wales, based on an independent Evidence Review Group (ERG) assessment of the manufacturer's submission. Since January 2007, the disease area with the most submissions has been oncology, accounting for 34% of the 243 interventions reviewed, followed by rheumatology and cardiovascular disease (CVD), each accounting for 11%. Overall, 20% of treatments were not recommended, with a disproportionately high proportion of cancer drugs being rejected (35% of 83 submissions vs. 4% of cardiovascular and 11% of rheumatology drugs).

Further investigation of the submissions since 2007 indicates that over 70% related to rheumatology, CVD, neurology and nephrology included NMA evidence, as opposed to under 40% of submissions in oncology, diabetes, orthopaedics, gastroenterology and respiratory medicine. It is also notable that, while there has been no obvious increase in overall proportion of submissions including an NMA (with this varying between one-third and two-thirds), the proportion of these analyses criticised by the ERG on methodological grounds has increased from under 60% in most years up to 2011 to over 80% of those published between January and May 2013. The forms of this criticism are described in the next section.

## SPECIFIC CRITICISMS BY NICE ERGs

The ERG criticisms can be classified as problems with the identification of data and study inclusion in the underlying systematic review; issues around the reporting of data; and concerns about the statistical analyses in the NMA.

### *1. Identification of data*

An NMA relies heavily on a systematic approach for identifying all relevant data on the efficacy of the comparator interventions. Criticisms made by the ERGs have reflected not only this requirement but also the need to search comprehensively across multiple databases and to be transparent in reporting the methodology.

Some criticisms of the search methodology concern issues that are unlikely to have substantially affected the results, but which reduce confidence in the rigour of the systematic review. These include the following:

- Failing to structure the search strategy clearly
- Failing to use a validated filter for RCTs
- Applying search engine limits incorrectly or not reporting them in enough detail

More serious problems, likely to have resulted in relevant studies being missed, include:

- Not searching all relevant sources
- Errors in search terms
- Omitting relevant search terms, or missing key publications
- Using an unsystematic process

Although failure to report the search strategy or methodology in adequate detail was a criticism for two submissions, this meant that the ERG could not verify the results of the analysis, rather than proving that the approach was methodologically unsound.

Manufacturers might be tempted to use existing systematic reviews to source data on competitors, to avoid costs and delay required to conduct a bespoke review. However, one manufacturer was criticised for using effect sizes from such a review rather than data from the primary RCTs. Also, three other submissions were criticised for basing their analysis on an out-of-date systematic review.

### *2. Study inclusion*

Inclusion criteria for any systematic review of efficacy and safety studies are focused on the PICOT criteria: the population, intervention, comparator, outcomes, and time points of interest. Accordingly, NICE's technology appraisal process starts with a scoping exercise, where key participants in the appraisal agree with the PICOT categories to be used

in the manufacturer's submission, so helping to ensure that the final guidance is based on evidence relevant to the target population. The ERG compares the submission to the agreed scope, so any deviation will probably be detected and criticised.

Various problems arise in seeking relevant evidence that allows a comparison between the new drug and the main comparators used in the UK at the time of the appraisal. Specifically, RCT data may not be available for these competitors in the specific subpopulation for whom the target drug is indicated. This has led to the following criticisms:

- The populations of included studies were broader than agreed in the scope
- Included populations were considered heterogeneous enough to make pooling of the data potentially misleading
- Unfounded assumptions were made about the similarity of treatment response across different subgroups
- The lack of reporting of baseline characteristics meant that the heterogeneity of the population could not be determined

In contrast, in another submission, the included population was considered too narrow.

For an indirect comparison to be valid, there have to be common treatment arms, in terms of the specific interventions and regimens that have been evaluated. This can make selection of appropriate comparators difficult, and it can be challenging to get the right balance. For example, submissions have been criticised for the following:

- Including too few relevant comparators (in rheumatology, CVD and cancer)
- Including too many comparators (in CVD)
- Including competitor agents unlicensed for the particular indication in the UK (in cancer)
- Failing to include unlicensed competitors commonly used for the condition (in cancer or ophthalmology disorders)

The selection of comparators, outcomes and time points for the analyses has also been criticised for these reasons:

- Including comparators differing from those in the agreed scope (in cancer and CVD)
- Analysing data from a comparator that was based on a dose or regimen that is not typically prescribed in the UK (in cancer, rheumatology and CVD)
- Not assessing the effect of the intervention on a key outcome (in rheumatology and respiratory disease)
- Not collecting outcome data for long enough (in CVD and rheumatology)
- Pooling data from studies that used different time points, and not adjusting the results accordingly (in cancer and CVD)
- Excluding studies for reporting data at different time points (in orthopaedics)

### *3. Reporting of NMAs*

Detailed reporting of the methodology is essential for assessing the rigour of an NMA. It is telling, therefore, that a featured ERG criticism related to insufficient information provided on one or more of the following aspects of studies included in or excluded from such analyses:

- Baseline characteristics among the included studies (in hepatitis)
- Interventions and comparators joined in a network (in CVD and cancer)
- Potential sources of heterogeneity (in CVD and mental health)
- Reasons for exclusion of studies (in respiratory disease)

This missing information meant that the ERG could not determine whether it was appropriate to include or exclude studies and therefore, subsequently join (or not join) the relevant interventions in a network, or whether there was significant heterogeneity among studies that would affect the results of the NMA. Other flaws that hindered assessment of the NMA's validity were unclear information on the source of some of the data used in the analysis, or, in one instance, on whether double data extraction and/or validation had been used to check the data reported.

Anomalies in the information criticised in several submissions included the following:

- Inconsistencies where the same data were, or should have been, reported in different sections of the submission
- Inconsistency between the number of trials included in the NMA and reported in the submission
- Errors in the data fed into the NMA, for example, reporting 12-week data as 6-week outcomes, or inappropriate use of median survival times

These errors cast raise doubts on the NMA's findings, and, in some cases, were substantial enough to change the direction of the effect size for the comparison.

Information on study quality is required to judge the overall quality of the NMA results and whether there are important differences between trials that would make pooling of their data inappropriate. In various submissions, a quality assessment was incomplete, incorrect or not reported.

### *4. Data analysis*

The most common criticism of data analysis in an NMA related to heterogeneity among the included studies. The ERGs were clear that studies showing significant heterogeneity should not be pooled in an NMA, and specifically cited

concerns about inappropriate pooling in the face of between-trial differences in population, intervention, comparators, time points, outcomes and study design.

Multiple submissions (in cancer, rheumatology, CVD, and virology) failed to assess the underlying heterogeneity across included studies and so could not confirm that it was appropriate to combine them in a network. Submissions that noted heterogeneity (also in cancer, CVD and rheumatology) were also criticised if this heterogeneity was not adjusted for or explored (e.g., using meta-regression or subgroup analysis).

Manufacturers were also criticised for their choice of statistical methods. ERGs criticised several submissions for an inappropriate or unclear rationale for the use of fixed- versus random-effects models. In general, where there was heterogeneity among the included trials, the ERGs preferred use of a random-effects model to take into account some of this variability. They also favoured the use of log hazards ratios, rather than a comparison of pooled treatment arm-level data such as median survival times derived from a meta-analysis. In the pooled analysis, any statistical benefit from randomisation is lost, and the data are only relevant to one time point, rather than incorporating information over the whole timescale as in a hazard ratio. Criticisms of methods used for statistical analysis have been levelled at submissions related to cancer, neurology, rheumatology and virology.

Of note, three submissions in cancer and ophthalmology were criticised for failing to conduct an NMA when the ERG considered such analysis feasible.

## IMPACT OF CRITICISMS OF NMAs ON NICE DECISION?

This overview has considered the types of criticisms made by ERGs of NMAs from manufacturers' submissions. However, some submissions received multiple criticisms, while others were faulted on just one or two points. For the one-fifth of submissions since 2007 not approved by NICE, the rejection usually related to weakness of the primary evidence supporting the clinical and cost-effectiveness of the new drug rather than to problems with the NMA.

Ongoing research will review the ERG criticisms in more detail to determine which, if any, avoidable flaws in the NMA process are associated with rejection of submissions, and how far assessment of the evidence on competitor products might help in designing clinical trials that facilitate such indirect comparison.

FOR THE ONE-FIFTH OF SUBMISSIONS SINCE 2007 NOT APPROVED BY NICE, THE REJECTION USUALLY RELATED TO WEAKNESS OF THE PRIMARY EVIDENCE SUPPORTING THE CLINICAL AND COST-EFFECTIVENESS OF THE NEW DRUG RATHER THAN TO PROBLEMS WITH THE NMA.

For more information, please contact Alison.Martin@evidera.com, Maria.Rizzo@evidera.com or Ike.Iheanacho@evidera.com.