Novel Methods for Indirect Comparison of Treatments: When Are They Needed and How Do They Work?

K. Jack Ishak, PhD, MSc, Executive Director, Biostatistics and Senior Research Leader, **Irina Proskorovsky, BSc**, Senior Biostatistician, and **Agnes Benedict, MSc, MA**, Senior Researcher, Evidera, St-Laurent, QC, Canada; and **Connie Chen, PharmD**, Senior Director, Global Outcomes Research, Pfizer Inc., New York, NY, USA

KEY POINTS . . .

Network meta-analyses (NMA) are the gold standard for indirect comparisons of treatments, but incomplete evidence networks and heterogeneity (among other things) between studies may limit the use of NMAs.

Simulated treatment comparisons (STCs) and matching adjusted indirect comparisons (MAIC) can overcome these challenges by carrying out a targeted comparison between outcomes for specific treatment arms of interest.

Statistical adjustment is required to reduce or minimize confounding in the comparisons. STCs achieve this with use of predictive equations, while MAIC relies on reweighting patients.

Indirect Comparison of Treatments

There is generally a paucity of evidence about the relative effectiveness of a new treatment compared with its competitors. And yet, this is a critical consideration in reimbursement decisions, as well as in the planning of future research. In the absence of head-to-head studies, comparative evidence is derived through indirect comparisons, relying on common comparators to link data from trials of the various treatments of interest. That is, treatments A and B, which were compared with treatment C in their respective trials, can be indirectly compared with each other by contrasting effects of A vs. C to that of B vs. C. Network meta-analysis (NMA) or mixed/indirect treatment comparison (MTC) [1-4] is the standard technique used for this purpose. These methods are broadly used and accepted by the research community as well as health technology assessment agencies, in part because they can incorporate data from all competing treatments in a therapeutic area, thus reflecting the totality of evidence that is available.

In some cases, however, NMAs may not be able to produce the comparisons of interest (i.e., when common comparators are not available), or may be subject to limitations (e.g., heterogeneity between trials) affecting the reliability of the results. A recent review of 243 individual assessments in 181 technology appraisals conducted by NICE between 2006-April 2013 examined from the point of view of NMAs [5] shows that over half of the assessments (54%) did not include an MTC. Of these, nearly a quarter (24%) cited heterogeneity as a challenge. Of those that included an NMA, 25% had not addressed heterogeneity.

Two alternative approaches—simulated treatment comparisons (STCs) [6,7] and matching adjusted indirect comparisons (MAICs) [8,9] can overcome these issues by making targeted comparisons of outcomes for a new treatment and those observed for its comparators using data from the treatment arms of their respective studies. Thus, the units of analysis in these targeted comparisons are outcome measures like event rates rather than relative effect estimates like hazard ratios as in MTCs. This poses an important challenge, however; outcomes in treatment arms from different studies are not necessarily comparable because observed differences may not be only attributable to the treatments, but may also reflect differences in the characteristics of the patients in the two studies as well as features of the study designs (e.g., blinding). Targeted comparisons are designed to deal with these issues and produce reliable comparisons by making analytical adjustments to balance the populations being compared. Unlike NMAs that rely only on published data, targeted comparisons require patient-level data on at least one of the treatments to be able to adjust for differences in populations.

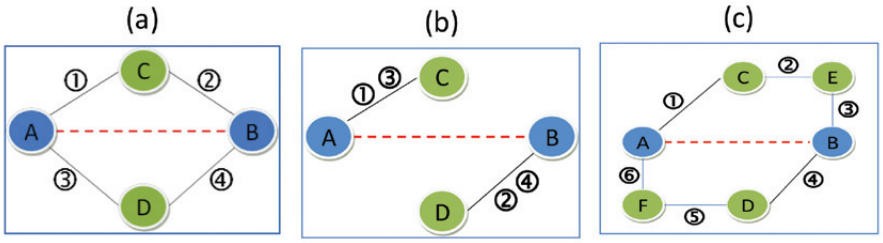
When Should Novel Approaches Be Considered?

STCs and MAICs can be applied, or at least considered and assessed for feasibility, in situations where standard techniques pose limitations or cannot be applied at all. Three specific scenarios are described below.

Heterogeneity

Figure 1a illustrates a simple evidence network (i.e., representation of the studies and treatments involved in the NMA) to evaluate a comparison of treatments A and B. The network includes 4 studies, identified by lines connecting the treatments compared in each of these. For instance, trial 1 compared treatment A to C, and trial 4 compared treatment B to D. Thus, the indirect comparison of A and B (represented by the dashed red line) is informed by the relative effects of these treatments to their common comparators C and D. That is, study 3 provides evidence about A vs. D, and, thus, links A to B since D was also the comparator in study 4. Similarly, study 2 compared treatments B and C, which allows A to be linked to B since study 1 also used treatment C as the comparator. Thus, the indirect comparison of A vs. B

Figure 1. Examples of possible evidence network based on four studies providing indirect data about the relative effectiveness of treatments A and B. Panel (a) represents a complete network; panel (b) represents an incomplete network; panel (c) represents a network requiring multiple intermediate comparisons.



would pool evidence from these two paths in the evidence network; the reliability of this depends on the assumption that all involved studies measure the same underlying effect (i.e., are homogeneous), or at most differ only in a random way.

Suppose, however, trials 2 and 3 have similar populations and design, and differ significantly from the other two studies. Such variation causes heterogeneity in the results being pooled and compared, which is dealt with in NMAs by adding parameters that account for excess variability in results. This assumes, however, that differences between trials only cause random fluctuation, so that the indirect comparison derived from the NMA effectively averages over differences in populations, design features, measurement techniques, etc., across studies. This can be problematic, however, when there is significant heterogeneity, and specific differences that may distort results can be identified. Published data are often too limited to allow a closer examination and adjustment for such factors in NMAs.

STCs and MAICs can deal with this type of heterogeneity by focusing the comparison of the studies that are deemed more closely comparable—2 and 3 in this example. Outcomes observed for treatment A in study 3 can be compared with outcomes for B in study 2. It is possible that the profiles of the populations of these studies may differ (even if only due to chance) and requires adjustment to obtain an unbiased comparison. The way this is handled in each approach is further described below.

Incomplete Evidence Network

STCs and MAICs would also be useful in situations where the evidence network is incomplete or disconnected. That is, the treatments to be compared cannot be linked through common comparators. This is illustrated in Figure 1b, where two trials comparing A to C and two trials comparing B to D make up the evidence network. Since the comparators in the trials of A and B are different, it is impossible to obtain an indirect comparison of these treatments with an NMA. Approaches like STC or MAIC may be the only way to achieve an indirect comparison in these situations, since the analyses would rely only on the outcomes from the arms receiving treatments A and B, thus bypassing the need for a completely linked network. Indeed, STC and MAIC can derive comparisons of treatments in single-arm trials, whereas these, by definition, cannot be incorporated in evidence networks since they lack comparator arms.

In the example illustrated in Figure 1b, two studies are available for both treatment A and B, leaving the analyst with a decision to make about which studies to use for the comparison. A natural choice may exist if a specific pair is more compatible because of similar populations and design. Otherwise, one may also apply STC or MAIC for all possible pairs of studies (i.e., 1 vs. 2, 1 vs. 4, 3 vs. 2 and 3 vs. 4), each of which would yield an indirect comparison of A vs. B. These results can then be pooled using traditional techniques for meta-analysis to obtain an average estimate.

Multi-Step Comparison

STCs and MAICs may also be useful in situations where the treatments of interest can only be linked through multiple intermediate comparisons. This is illustrated in Figure 1c. In this evidence network, trials of A and B do not have a common comparator, and must rely on trials that compared their respective comparators to make the link. That is, A is linked to B through a comparison of C to E and F and D (i.e., A vs. C, C vs. E, E vs. B, and A vs. F, F vs. D and D vs. B). The reliability of MTCs in this situation may be compromised as heterogeneity may impact comparisons at intermediate steps and distort the main comparison of interest, which can lead to large uncertainty in estimates of relative effects. The problem is amplified as the number of steps involved to link treatments increases (e.g., to link A to D in Figure 1c). The targeted comparisons involved in STCs and MAICs bypass the issue by directly comparing outcomes in the specific arms of interest after adjustment for baseline differences, as long as the trials of treatment A and B can be considered sufficiently compatible.

When Are Novel Approaches Feasible?

The first consideration in assessing the feasibility of STC or MAIC lies in the availability of patient-level data on at least one of the treatments being compared. This should be possible when analyses are initiated by the manufacturer of one of the treatments being compared. One or more trials of the manufacturer's product (the index trial(s)) would then serve as the basis of the STC and MAIC and would be used to adjust for differences in populations of comparators' trials. In most situations, data on the comparator treatments will only be available from publications. This is not a limiting factor, as long as information on the profile of the population and outcomes of interest are reported with adequate detail.

Availability of data from index and comparator studies makes analyses computationally feasible, but does not necessarily mean results will be reliable. Ultimately, this is determined by whether the index and comparator studies are reasonably compatible. Compatibility is determined based on the similarity of the populations and the designs of the trials. It is not necessary for the populations of the index and comparator trials to be identical, since the methods are designed specifically to balance differences. This can only be done, however, when there is sufficient overlap in the profiles of the two samples. For example, suppose gender is an important determinant of outcomes, so that differences between studies in the mix of male and female patients can confound comparisons. STC and MAIC can adjust for differences in the proportion of male vs. female patients, but the comparison may be unreliable if one study was based mostly on male patients and the other >

mostly on female patients. Indeed, the adjustment would be impossible if one study only included males while the other included only female patients. Thus, some reasonable overlap is required between the distributions of baseline characteristics across the studies to have reliable adjustment. Similarly, the duration of the trials and timing of measurements should be similar but not necessarily identical; for instance, a comparison of survival between studies where follow-up lasted 5 vs. 3 years is appropriate, but may be misleading if the gap was 5 vs. 1 year; likewise for other design features such as admissibility criteria, concomitant medications, treatment protocols, etc. Details about these features must be reviewed carefully to assess whether they overlap enough to assume that results from the each of the studies are applicable in the setting of the other. This may require clinical insight and opinion and is not strictly a statistical issue.

Finally, reliable application of STCs and MAICs requires that all determinants of the outcomes of interest that may confound the comparison are available in both the index trial data and reported in the publication(s) for the comparator(s) (which will be in summary form, such as means and percentages). The results are subject to residual confounding in cases where determinants are available in one but not both sources. The impact of missing determinants or other differences in study designs (sometimes referred to as "study effects") may be assessed when the index and comparator studies used the same treatment (or placebo) in their reference arms by applying STC or MAIC to compare outcomes in the two. Ideally, when studies are deemed compatible, study effects should be minimal, adjusting for differences in the population characteristics. That is, the STC or MAIC of the reference arms would yield no or very small difference in outcomes in the absence of residual confounding or study effects. Large differences between outcomes in the common reference arms (e.g., of similar magnitude to effects produced in the indirect treatment comparison) would suggest possible bias in the analyses, and should be considered carefully to understand their potential causes and implications for the suitability of the comparison. Results for the index and comparator treatment might have to be further adjusted by the magnitude of the residual difference between reference arms to get a more reliable estimate.

How Do Targeted Comparisons Work?

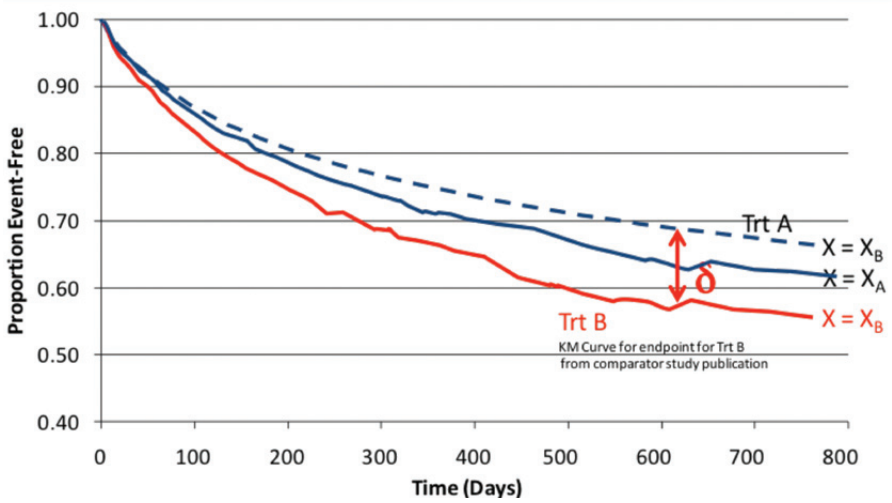
STCs and MAICs are very similar conceptually. Figure 2 shows a representation of how balanced comparisons are derived in STCs and MAICs. In this illustration, the outcome of interest is a time-to-event endpoint. The solid blue line represents the time-to-event distribution from the index trial of treatment A, while the red solid line represents the distribution for the comparator B obtained from a published report or manuscript. A comparison of these lines is biased by the fact that the profile of the population represented in the blue line (denoted by XA) may differ, even if only by chance to the profile in the red line (XB). Thus, to adjust for potential imbalances, these methods aim to generate an adjusted time-to-event curve that reflects what outcomes may have been with treatment A in a population that matches the profile for treatment B. This is represented by the dashed blue line, which can now be compared directly with the observed outcomes for treatment B (i.e., red line) to measure the relative effectiveness of A and B (denoted by TRT).

STCs and MAICs differ in the way they generate the adjusted outcomes for treatment A (dashed blue line). STC accomplishes this by creating a predictive equation for each outcome being compared. The equations are then used to predict outcomes that would have been observed for treatment A in patients with characteristics matching those in XB. That is, the adjusted line is produced by setting predictors to their corresponding values in XB.

MAICs deal with the adjustment by reweighting patients in the index trial so that the weighted average values of determinants of outcomes in the index trial (i.e., XA) match XB. These weights are derived from a propensity-score-type analysis using the index trial data, predicting membership into the index vs. comparator's trial. Thus, unlike in STC where equations predict outcomes, the propensity equation in MAIC predicts membership to the index vs. comparator trial. An individual weight is then predicted for each patient in the index trial and applied in Kaplan-Meier analyses (for example) to generate the adjusted curve.

In both STC and MAIC, the adjusted outcomes obtained for the index treatment reflect how this treatment would have performed in a population that matches the comparator population, and can be compared with observed results for the latter. This produces an estimate of the relative effectiveness along with measures of uncertainty, like standard errors or confidence intervals. The measure used to capture treatment differences depend on the type of outcome being considered (e.g., differences in means for continuous outcomes, odds ratios for dichotomous outcomes, etc.).

Figure 2. Conceptual representation of how comparisons are derived with adjustment for differences in population profiles with STC and MAIC.



KM, Kaplar Maier; MAIC, matching adjusted indirect comparison; STC, simulated treatment comparisons; Trt, treatment.

When to Choose Simulation Versus Matching-Based Approach?

STCs and MAICs are conceptually very similar and use the same data to accomplish the goal of adjustment for potential confounding. It is, therefore, reasonable to expect that the two methods would yield similar results. Some differences between these methods are worth noting, however.

STCs involve generating predictive equations for each of the outcomes of interest. Deriving these equations provides added insight, by showing the relative predictive strength (and, hence, influence) of each predictor involved in the adjustment for confounding. In some situations, the use of equations can provide greater flexibility than is possible with MAIC. For instance, when one or more characteristics differ significantly between the studies, MAICs may have to assign large weights to a small portion of the index population to balance the groups. In fact, in some cases, matching a characteristic may not be possible when their distribution in the two studies differ substantially. In STC, the effect of the characteristic is modeled through the equation, which can be used to then extrapolate beyond the range in the index trial (within some reasonable, clinically appropriate margin).

Some differences between STCs and MAICs lie in potential efficiencies associated with each approach. STCs can be more efficient than MAICs in situations where comparisons with multiple comparators are to be made for a small set of outcomes. Equations for the outcomes would be derived once from the index trial and applied with data from each comparator treatment's study. With MAIC, a separate set of weights would be required for each comparator treatment's study population. By the same token, MAICs would offer efficiencies in situations where there is a single comparator but many outcomes to be compared. A single set of weights would be required to balance the two populations, and could be applied in analyses for each outcome.

Strengths and Limitations

STCs and MAICs are robust and reliable methods to derive indirect comparisons between treatments. These novel methods can produce comparative evidence in situations where standard techniques are inadequate, but can also be complimentary to NMA, providing a more targeted assessment of the relative effectiveness of the treatments. Whereas the NMA may provide an averaged effect estimate, by using the index trial as the basis of the analysis, the STC or MAIC reflects the relative effectiveness that might have been observed if the comparator had been included as an additional arm in the index trial.

It is important to note that STC and MAIC rely on important assumptions. The studies included in the comparison must be compatible, such that the effects observed from one can be deemed applicable to the population of the other trial. Any differences in design or background factors must be negligible or possible to adjust for analytically. All determinants of outcomes of interest must be measured to fully account for possible confounding. The reliability of the analyses depends on the successful matching of populations in MAIC and the accuracy of the equations in STC. The distribution of weights assigned to patients in the index trial in MAIC must be carefully assessed to ensure that these do not give large influence to specific subgroups or individuals. In STC, missing predictors and incorrect shape of the equation can distort predictions and bias comparisons. With both methods, effect modification and subgroup effects cannot be captured, since published data are not typically available with sufficient detail in subsets of the population.

These novel methods can produce comparative evidence in situations where standard techniques are inadequate, but can also be complimentary to NMA, providing a more targeted assessment of the relative effectiveness of the treatments.

References

[1] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;50:683-91. [2] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897-900. [3] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105-24. [4] Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21:2313-24. [5] Martin A, Rizzo M, Iheanacho I. Faulty Connections: Can criticisms of network meta-analysis in nice submissions be avoided? Value Health 2013;16:A608. [6] Caro JJ, Ishak KJ. No head-to-head trial? simulate the missing arms. Pharmacoeconomics 2010;28:957-67. [7] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. Pharmacoeconomics 2015;33:537-49. [8] Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health 2012;15:940-7. [9] Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics 2010;28:935-45. ■