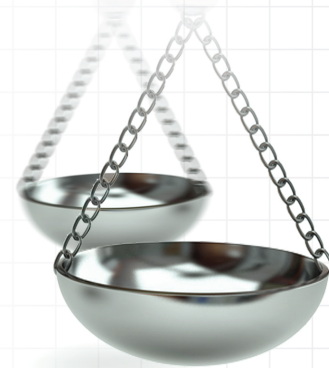




Number needed to treat in indirect treatment comparison

Patricia Guyot^{*1}, Wei Cheng^{1,2}, Gabriel Tremblay³, Ronda Copher³, Heather Burnett⁴, Xuan Li³ & Charles Makin⁵

¹Real World Strategy & Analytics, Mapi Group, 27 rue de la Villette, Lyon 69003, France

²The Knowledge Synthesis Group, Ottawa Hospital Research Institute, 501 Smyth Road, PO Box 201B, Ottawa, Ontario K1H 8L6, Canada

³Eisai, 155 Tice Boulevard, Woodcliff Lake, NJ 07677, USA

⁴Real World Strategy & Analytics, Mapi Group, 40 Court Street, Suite 410, Boston, MA 02108, USA

⁵RWE & Late Phase Research, ICON plc, 2100 Pennbrook Pkwy, North Wales, PA 19454, USA

* Author for correspondence: Tel.: +33 47 213 5146; Fax: +33 617 722 4096; pguyot@mapigroup.com

Journal of **Comparative Effectiveness Research**

Aim: For dichotomous outcomes, odds ratio (OR) is one of the usual summary measures of indirect treatment comparison. A corresponding number needed to treat (NNT) estimate may facilitate understanding of the treatment effect. **Methods:** We show how to estimate NNT based on OR results of a matching adjusted indirect comparison. We also have derived the explicit formula of its 95% CIs by applying the delta method, and as an alternative, a simulation-based method. **Results:** The method was applied in a case study example in radioiodine-refractory differentiated thyroid cancer (RR-DTC) patients, comparing lenvatinib to sorafenib. For every two RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have progressed and for every eight RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have died. **Conclusion:** Using NNT to summarize the results of a matching adjusted indirect comparison can help the clinicians to better understand the results in addition to OR.

First draft submitted: 5 April 2017; Accepted for publication: 9 October 2017; Published online: 14 March 2018

Keywords: evidence synthesis • indirect treatment comparison • number needed to treat

To treat or manage diseases, new interventions are routinely developed as improvements of existing therapies. To assess the potential improvement of therapies, decision makers, such as patients, providers or payers, require information on the relative or comparative effectiveness of the treatments [1–3]. Typically, clinical outcomes measured in randomized clinical trials (RCT) provide an established means of demonstrating treatment efficacy compared with placebo or standard of care. Additional evidence related to side effects and patient-reported outcomes are also collected and used to demonstrate the value of a treatment. RCTs are considered the gold standard for obtaining the most reliable evidence [4]. If more than one RCT is conducted on the same two interventions, the evidence can be summarized via pairwise, conventional meta-analysis. Though ideal from an evidence generation perspective, given resource limitations, new treatments are not directly compared with all other existing treatments. In the absence of direct treatment comparisons, methodological techniques offer useful options to evaluate treatments relative to other therapies.

Indirect treatment comparison (ITC) allows for multiple pairwise comparisons across a range of treatments to be compared in the absence of direct evidence [5–7]. Typically, if one RCT compares treatment A versus placebo and a second RCT compares treatment B versus placebo, the comparative efficacy of A versus B can be estimated by subtracting the two relative treatment effects using an appropriate scale (e.g., log odds ratio [OR] for dichotomous outcomes). The comparison between treatments A and B is made via a common comparator (usually placebo). When all the direct evidence is graphically displayed, a network diagram is obtained (Figure 1). Each node represents a treatment, and the edges indicate the direct comparisons available. When all active treatments are compared only versus a common comparator, the analysis is called ITC. If evidence is available from both direct and indirect comparisons, the term network meta-analysis (NMA) is usually used.

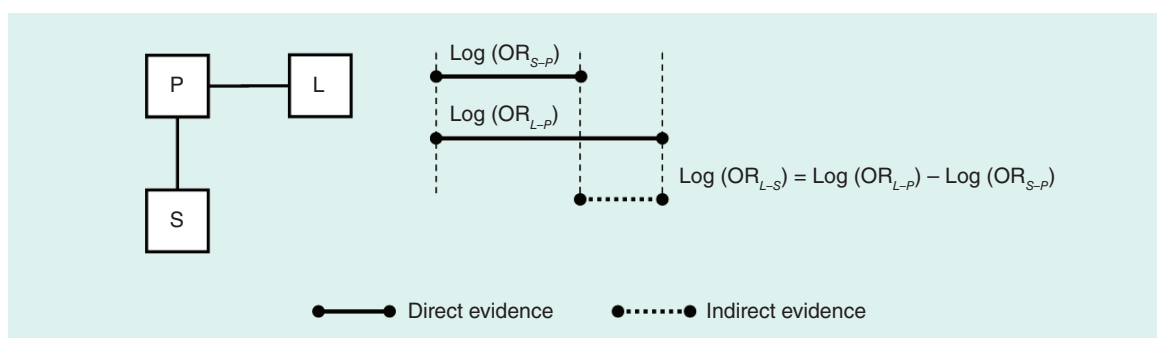


Figure 1. Indirect treatment comparison of L versus S based on one L-placebo and one S-placebo trial assuming the consistency equation.
OR: Odds ratio.

Patient characteristics are not necessarily comparable in pairwise meta-analyses, NMA and ITC. The benefit of randomization to treatment holds within a given RCT but not between them. If patient characteristics are not similar across the RCT-leveraged studies, the transitivity in the consistency assumption [8] is violated and results of the ITC are subject to bias. Treatment effect modifiers that could affect patient comparisons across RCTs can be part of the study design, in other words, inclusion/exclusion criteria or other patients' characteristics. The definition of the outcome measured might also vary across the RCTs and influence the treatment effect. For example, if the study of treatment A versus placebo is conducted with only patients with stage II breast cancer and the study of treatment B versus placebo evaluated therapy in patients with stage IV disease, the indirect relative overall survival (OS) estimate of treatments A versus B is not based on a comparable population, and thus is prone to bias.

In the presence of treatment effect modifiers, researchers can create or rebalance the population of two trials using approaches such as a matching adjusted indirect treatment comparison (MAIC) [9,10]. An established means of implementing MAIC uses available individual patient data (IPD) to match the summary patient characteristics reported in RCT without IPD. The outcomes of interest are then compared across balanced study populations, in other words, with similar summary treatment effect modifiers characteristics. The matching is achieved by reweighting patients in the trials with available IPD, so that weighted mean baseline characteristics matched those reported for the study without IPD [10]. The more comparable the populations of the trials are, the more likely the validity of the consistency assumption holds. Even after adjustments, there is still a risk that the two trials populations differ, due to missing or unmeasured treatment effect modifiers.

There are many ways to describe a treatment effect for binary outcomes: OR, relative risks/risk ratio, absolute risk reductions (ARR)/risk difference and numbers needed to treat (NNT) can be reported. ORs are computed as the ratio of two odds (i.e., the odds that an event will occur, divided by the odds that it will not occur) and are often used to present the relative efficacy of treatments derived through meta-analysis [11,12]. The relative effects of binary outcomes from each study are often modeled on the log-odds scale; if not, they can be converted to the log-odds scale, therefore throughout this article, we only present how to transform OR into NNT. NNTs offer one advantage in addition to ORs, in that they present a relatively straightforward clinical interpretations [13,14], providing that the population and outcome on which the analyses were based is clearly defined.

NNT corresponds to the inverse of an ARR [15,16], which is computed as the difference between event rates in two groups (the difference between the expected event rate [EER] and control event rate [CER]). NNT indicates the number of patients who need to be given a new treatment instead of the comparator to have one more patient benefit. When an NNT reflects an undesirable event (e.g., adverse effect), it is referred to as the number needed to harm. Dias *et al.* [17] have described methods to obtain NNT with 95% CIs based on the results of a Bayesian NMA. The scenario in this article is the ITC/MAIC of only two active treatments with a common comparator, not NMA, and should not be compared with Bayesian NMA. For simple ITC of relatively few treatments with a common comparator, the frequentist approach is well accepted by authorities (e.g., Canadian Agency for Drugs & Technologies in Health [3] and Pharmaceutical Benefits Advisory Committee [1]). MAICs are typically performed using a frequentist approach for simplicity. In this article, we present a simple process to obtain an NNT estimate with 95% CIs using results from frequentist ITC/MAIC.

To demonstrate the validity and usefulness of this approach, a case study comparing two novel treatments for radioiodine-refractory differentiated thyroid cancer (RR-DTC) is presented. Two treatments, lenvatinib and sorafenib are currently licensed in Europe and the USA to treat RR-DTC. No head-to-head clinical studies comparing the two active treatments have been carried out in this patient population. However, both treatments were compared versus placebo in two individual trials – the SELECT study for lenvatinib [18] and the DECISION study for sorafenib [19]. Lenvatinib is indirectly compared with sorafenib in this paper and the summary results expressed in terms of NNT.

Materials & methods

Meta-analysis used directly on ARR usually displays more heterogeneity than meta-analysis on OR [20,21]. Given this limitation, the Cochrane Handbook for Systematic Reviews of Interventions advises to first perform evidence synthesis on OR and then to derive ARR and NNT from it. A robust technique to obtain NNT based on results from ITC or MAIC therefore consists of first obtaining the indirect OR estimates by subtracting the log ORs, and then deriving the corresponding indirect ARR and NNT estimates, using the formula from Sackett *et al.* [22] (also cited in Furukawa *et al.* [23]).

To estimate 95% CIs around the ARR and often NNT, the Wald method is regularly used for one specific trial [24], which is not applicable across trials. We propose two approaches to estimate the 95% CI of NNT based on results from MAIC: the delta method [25] or the simulation-based method.

OR results are often difficult to understand and translate into clinical practice. Although the trend of the effect is easy to understand, the magnitude is less clear. NNT provides a more approachable way to clinically represent the treatment efficacy, with results being easier to communicate to a broader audience.

For two trials A and B, with trial A comparing treatment L to placebo and trial B comparing treatment S to placebo, we define the following notations:

- n_L : number of events in the L arm of the A trial
- N_L : total number of patients in the L arm of the A trial
- n_{p1} : number of events in the placebo arm of the A trial
- N_{p1} : total number of patients in the placebo arm of the A trial
- n_S : number of events in the S arm of the B trial
- N_S : total number of patients in the S arm of the B trial
- n_{p2} : number of events in the placebo arm of the B trial
- N_{p2} : total number of patients in the placebo arm of the B trial

The estimator of the EER [22] for treatment L, $\widehat{EER}_L$, is then $\widehat{EER}_L = n_L / N_L$, and likewise $\widehat{EER}_S = n_S / N_S$ for treatment S. The CER [22] for the two trials are $\widehat{CER}_{p1} = n_{p1} / N_{p1}$ and $\widehat{CER}_{p2} = n_{p2} / N_{p2}$, respectively.

It should be noted that in the absence of a head-to-head trial comparing L versus S, the validity of the consistency assumption cannot be tested. As such, the following methods rely on the comparability of the included RCTs (Figure 1). The log (OR_{L-S}), is assumed equal to the log OR of L versus placebo in the A trial, log (OR_{L-p1}), minus the log OR of S versus placebo in the B trial, log (OR_{S-p2}):

$$\log(OR_{L-S}) = \log(OR_{L-p1}) - \log(OR_{S-p2})$$

The OR estimator of L versus S is therefore obtained using Equation 1 below:

$$\widehat{OR}_{L-S} = \widehat{OR}_{L-p1} / \widehat{OR}_{S-p2} = \frac{n_L}{N_L - n_L} \times \frac{N_{p2} - n_{p2}}{n_{p2}} \times \frac{N_S - n_S}{n_S} \times \frac{n_{p1}}{N_{p1} - n_{p1}} \quad (\text{Equation 1})$$

The NNT estimate of L versus S, $\widehat{\text{NNT}}_{L-S}$, is then obtained according to the formula from Sackett *et al.* [22], which is written in Equation 2:

$$\widehat{\text{NNT}}_{L-S} = \frac{1 - \widehat{\text{EER}}_S + \widehat{\text{OR}}_{L-S} \times \widehat{\text{EER}}_S}{(1 - \widehat{\text{EER}}_S) \times \widehat{\text{EER}}_S \times (1 - \widehat{\text{OR}}_{L-S})} \quad (\text{Equation 2})$$

this is algebraically identical to the reciprocal of the estimator of ARR:

$$\widehat{\text{NNT}}_{L-S} = \frac{1}{\widehat{\text{ARR}}_{L-S}} \text{ and } \widehat{\text{ARR}}_{L-S} = \widehat{\text{EER}}_S - \frac{\widehat{\text{OR}}_{L-S} \times \widehat{\text{EER}}_S}{1 - \widehat{\text{EER}}_S + \widehat{\text{OR}}_{L-S} \times \widehat{\text{EER}}_S}$$

Regarding the 95% CI for NNT, it can be obtained by inverting the upper and lower boundaries of the 95% CI of ARR [11,18]. The 95% CI around ARR can be obtained as follows:

$$95\% \text{ CI } (\widehat{\text{ARR}}_{L-S}) = \left[\widehat{\text{ARR}}_{L-S} - 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}), \widehat{\text{ARR}}_{L-S} + 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}) \right]$$

with $\widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S})$ the standard error estimator for ARR of L versus S.

Within a trial and for large sample sizes and risks not close to zero or one, the variance around ARR for S versus placebo (or L vs placebo) can be estimated using the usual Wald method [26]:

$$\widehat{\text{var}}(\widehat{\text{ARR}}_{S-P_2}) = \frac{\widehat{\text{CER}}_{P_2}(1 - \widehat{\text{CER}}_{P_2})}{N_{P_2}} + \frac{\widehat{\text{EER}}_S(1 - \widehat{\text{EER}}_S)}{N_S}$$

However, this equation cannot be used to calculate the variance around ARR for L versus S, which is based on the MAIC. Instead, two approaches were developed to obtain the 95% CI for the NNT of L versus S.

Considering $\widehat{\text{ARR}}_{L-S}$ as a function of $\widehat{\text{OR}}_{L-S}$ and $\widehat{\text{EER}}_S$, the standard error estimator of $\widehat{\text{ARR}}_{L-S}$ can be derived explicitly using the delta method. Alternatively, one can generate $\widehat{\text{OR}}_{L-S}^{\text{sim}}$ and $\widehat{\text{EER}}_S^{\text{sim}}$ and the implicit correlation between them by simulation, and calculate the NNT as the reciprocal of the mean of simulated $\widehat{\text{ARR}}_{L-S}^{\text{sim}}$.

The 95% CI using the delta method

The 95% CI of $\widehat{\text{NNT}}_{L-S}$ is 'obtained simply by taking reciprocals of the values defining the 95% CI for the ARR' (Altman [27] and Bender [28]):

$$\left(\frac{1}{\widehat{\text{ARR}}_{L-S} + 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S})}, \frac{1}{\widehat{\text{ARR}}_{L-S} - 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S})} \right)$$

if the 95% CI of $\widehat{\text{ARR}}_{L-S}$ does not cover zero, or

$$\left(-\infty, \frac{1}{\widehat{\text{ARR}}_{L-S} - 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S})} \right) \cup \left(\frac{1}{\widehat{\text{ARR}}_{L-S} + 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S})}, \infty \right)$$

if the 95% CI of $\widehat{\text{ARR}}_{L-S}$ encompasses zero, in other words, the lower bound $\widehat{\text{ARR}}_{L-S} - 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}) < 0$ and the upper bound $\widehat{\text{ARR}}_{L-S} + 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}) > 0$. The negative part pertains to NNT (benefit), the esti-

mated number of patients who need to be treated with lenvatinib rather than sorafenib for one additional patient to benefit, and the positive part pertains to NNT (harm).

When an estimator is not a simple sum of observations, deriving its variance is not straightforward. The delta method is a procedure that is useful in that case. The basic idea is to use Taylor series expansion to derive a linear function that approximates the more complicated function.

ARR_{L-S} is a function of two random variables, $\log \widehat{EER}_S$ and $\log \widehat{OR}_{L-S}$.

Using the delta method, we can derive the variance of $\widehat{ARR}_{L-S}$ as $\text{var}(\widehat{ARR}_{L-S}) =$

$$\begin{pmatrix} \frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{EER}_S} \\ \frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{OR}_{L-S}} \end{pmatrix}' \begin{pmatrix} \widehat{\text{var}}(\log \widehat{EER}_S) & \widehat{\text{cov}}(\log \widehat{EER}_S, \log \widehat{OR}_{L-S}) \\ \widehat{\text{cov}}(\log \widehat{EER}_S, \log \widehat{OR}_{L-S}) & \widehat{\text{var}}(\log \widehat{OR}_{L-S}) \end{pmatrix} \begin{pmatrix} \frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{EER}_S} \\ \frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{OR}_{L-S}} \end{pmatrix}$$

evaluated at $\widehat{OR}_{L-S}$ and $\widehat{EER}_S$, where the partial derivatives

$$\frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{EER}_S} = \frac{\partial \widehat{ARR}_{L-S}}{\partial \widehat{EER}_S} \frac{\partial \widehat{EER}_S}{\partial \log \widehat{EER}_S} = \left(1 + \frac{\widehat{OR}_{L-S}}{((\widehat{OR}_{L-S} - 1)\widehat{EER}_S + 1)^2} \right) \widehat{EER}_S (1 - \widehat{EER}_S)$$

$$\frac{\partial \widehat{ARR}_{L-S}}{\partial \log \widehat{OR}_{L-S}} = \frac{\partial \widehat{ARR}_{L-S}}{\partial \widehat{OR}_{L-S}} \frac{\partial \widehat{OR}_{L-S}}{\partial \log \widehat{OR}_{L-S}} = \frac{\widehat{EER}_S (1 - \widehat{EER}_S) \widehat{OR}_{L-S}}{(1 - \widehat{EER}_S + \widehat{OR}_{L-S} \times \widehat{EER}_S)^2}$$

and the elements in the variance-covariance matrix

$$\begin{aligned} \widehat{\text{var}}(\log \widehat{EER}_S) &= \left(\frac{\partial \log \widehat{EER}_S}{\partial \widehat{EER}_S} \right)^2 \widehat{\text{var}}(\widehat{EER}_S) \\ &= \frac{1}{\widehat{EER}_S^2 (1 - \widehat{EER}_S)^2} \frac{\widehat{EER}_S (1 - \widehat{EER}_S)}{N_S} \\ &= \frac{1}{N_S \widehat{EER}_S (1 - \widehat{EER}_S)} = \frac{N_S}{n_S (N_S - n_S)} \end{aligned}$$

$$\begin{aligned} \widehat{\text{var}}(\log \widehat{OR}_{L-S}) &= \widehat{\text{var}}(\log \widehat{OR}_{L-P_2}) + \widehat{\text{var}}(\log \widehat{OR}_{S-P_1}) \\ &\approx \frac{1}{n_L} + \frac{1}{N_L - n_L} + \frac{1}{n_{P_2}} + \frac{1}{N_{P_2} - n_{P_2}} + \frac{1}{n_S} + \frac{1}{N_S - n_S} + \frac{1}{n_{P_1}} + \frac{1}{N_{P_1} - n_{P_1}} \end{aligned}$$

$$\widehat{\text{cov}}(\log \widehat{EER}_S, \log \widehat{OR}_{L-S}) = -\widehat{\text{var}}(\log \widehat{EER}_S) \text{ because}$$

$$\begin{aligned}
 \text{cov}(\text{logit EER}_S, \text{log OR}_{L-S}) &= \text{cov}(\text{logit EER}_S, \text{log OR}_{L-P2} - \text{log OR}_{S-P1}) \\
 &= -\text{cov}(\text{logit EER}_S, \text{log OR}_{S-P1}) \\
 &= -\text{cov}(\text{logit EER}_S, \text{logit EER}_S) \\
 &= -\text{var}(\text{logit EER}_S)
 \end{aligned}$$

The 95% CIs using the simulation-based method

Assuming logit transformed EER of the sorafenib is normally distributed, we can simulate it with the mean and variance estimates $\widehat{\text{logit EER}}_S$ and $\widehat{\text{var}}(\text{logit EER}_S)$:

$$\text{logit}(\text{EER}_S)^{\text{sim}} \sim N(\widehat{\text{logit EER}}_S, \widehat{\text{var}}(\text{logit EER}_S))$$

$$\text{EER}_S^{\text{sim}} = \frac{\exp(\text{logit}(\text{EER}_S)^{\text{sim}})}{1 + \exp(\text{logit}(\text{EER}_S)^{\text{sim}})}$$

Analogously, simulate $\text{logit}(\text{EER}_L)^{\text{sim}}$ from the lenvatinib arm, $\text{logit}(\text{EER}_{p2})^{\text{sim}}$ from the placebo arm of the lenvatinib-placebo trial, $\text{logit}(\text{EER}_{p1})^{\text{sim}}$ from the placebo arm of the sorafenib-placebo trial,

$$\text{OR}_{L-S}^{\text{sim}} = \text{OR}_{L-P2}^{\text{sim}} / \text{OR}_{S-P1}^{\text{sim}}$$

$$= \left(\frac{\text{EER}_L^{\text{sim}}}{1 - \text{EER}_L^{\text{sim}}} \times \frac{\text{EER}_{p1}^{\text{sim}}}{1 - \text{EER}_{p1}^{\text{sim}}} \right) / \left(\frac{\text{EER}_S^{\text{sim}}}{1 - \text{EER}_S^{\text{sim}}} \times \frac{\text{EER}_{p2}^{\text{sim}}}{1 - \text{EER}_{p2}^{\text{sim}}} \right)$$

$$= \frac{\exp(\text{logit}(\text{EER}_L)^{\text{sim}}) \exp(\text{logit}(\text{EER}_{p1})^{\text{sim}})}{\exp(\text{logit}(\text{EER}_S)^{\text{sim}}) \exp(\text{logit}(\text{EER}_{p2})^{\text{sim}})}$$

$$\text{ARR}_{L-S}^{\text{sim}} = \text{EER}_S^{\text{sim}} - \frac{\text{OR}_{L-S}^{\text{sim}} \times \text{EER}_S^{\text{sim}}}{1 - \text{EER}_S^{\text{sim}} + \text{OR}_{L-S}^{\text{sim}} \times \text{EER}_S^{\text{sim}}}$$

Then we can calculate the mean of $\text{ARR}_{L-S}^{\text{sim}}$, denoted as $\overline{\text{ARR}_{L-S}^{\text{sim}}}$. We used the 97.5 and 2.5% quantiles of the simulated $\text{ARR}_{L-S}^{\text{sim}}$ to obtain the 95% CI. The NNT estimate is $1/\sqrt{\text{ARR}_{L-S}^{\text{sim}}}$ and the corresponding 95% CI of the NNT estimate is the reciprocal of the 97.5 and 2.5% quantiles of $\text{ARR}_{L-S}^{\text{sim}}$ if they are of the same sign.

If the 95% CI of $\widehat{\text{ARR}}_{L-S}$ encompasses zero, in other words, the lower bound $\widehat{\text{ARR}}_{L-S} - 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}) < 0$ and the upper bound $\widehat{\text{ARR}}_{L-S} + 1.96 \widehat{\text{SE}}(\widehat{\text{ARR}}_{L-S}) > 0$, then the negative part pertains to NNT (the estimated number of patients who need to be treated with L rather than S for one additional patient to benefit), and the positive part pertains to NNH (the estimated number of patients who need to be treated with L rather than S for one additional patient to be harmed).

The technique described in this article is fully reproducible. Equation 1 gives the estimation of an indirect OR; Equation 2 is the derivation of NNT estimates and two methods to estimate the 95% CIs around ARR and NNT (the delta method and the simulation based method) are provided in details.

Table 1. Study level data for progression-free survival, overall survival and objective response rate at 24 months.

Outcome	ACTIVE TX		Placebo		ARR (95% CI)	NNT (95% CI)
	EER	Total	CER	Total		
PFS						
SELECT (lenvatinib)	0.443	261	0.038	131	0.405 (0.336–0.474)	2.47 (2.11–2.97)
DECISION (sorafenib)	0.174	207	0.083	210	0.092 (0.028–0.155)	10.93 (6.45–35.91)
OS						
SELECT (lenvatinib)	0.728	261	0.641	131	0.087 (-0.011–0.185)	11.49 [†] (-∞ to -88.50) ∪ (5.40 ± ∞)
DECISION (sorafenib)	0.681	207	0.657	210	0.024 (-0.066–0.114)	41.67 [†] (-∞ to -15.08) ∪ (8.75 ± ∞)
ORR						
SELECT (lenvatinib)	0.648	261	0.015	131	0.633 (0.571–0.695)	1.58 (1.44–1.75)
DECISION (sorafenib)	0.122	196	0.005	210	0.117 (0.071–0.164)	8.51 (6.08–14.17)

[†]When the 95% CI of ARR encompasses zero, the 95% CI of NNT is in two pieces.
ARR: Absolute risk reduction; CER: Control event rate; EER: Expected event rate; NNT: Number needed-to-treat; ORR: Objective response rate; OS: Overall survival; PFS: Progression-free survival.

[†]When the 95% CI of ARR encompasses zero, the 95% CI of NNT is in two pieces.

ARR: Absolute risk reduction; CER: Control event rate; EER: Expected event rate; NNT: Number needed-to-treat; ORR: Objective response rate; OS: Overall survival; PFS: Progression-free survival.

Case study: MAIC of lenvatinib versus sorafenib

The method described above is applied to an indirect comparison of lenvatinib versus sorafenib. Both treatments were compared with placebo within separate RCTs.

The SELECT study was a 2:1 randomized, placebo-controlled, double-blind and multicenter international trial in subjects with progressive RR-DTC [18,19]. The RCT evaluated the efficacy and safety of lenvatinib 24 mg once daily versus placebo. The DECISION trial was a 1:1 randomized, placebo-controlled, double-blind, multicenter international trial in subjects with locally advanced or metastatic RR-DTC. This RCT evaluated the efficacy and safety of sorafenib 400 mg twice daily versus placebo [29].

Using data from both trials, three outcomes of interest were analyzed: progression-free survival (PFS), OS and objective response rate (ORR), both at 24 months. The EER, CER, ARR and NNT were estimated for each trial and each outcome. The results are presented in Table 1. From the within trial data both lenvatinib and sorafenib showed statistically significant NNT for PFS, but not OS versus placebo. At 24 months, for every 2.5 RR-DTC patients treated with lenvatinib and 10.9 patients treated with sorafenib, one patient would maintain stable disease while they would have progressed with placebo treatment. At 24 months, for every 11.5 patients treated with lenvatinib and 41.7 patients treated with sorafenib, one patient would survive while they would have died in the placebo arm. At 24 months, for every 1.6 patients treated with lenvatinib and 8.6 patients treated with sorafenib, one would have shown response to treatment while they would not have responded to placebo.

There were differences in the inclusion/exclusion criteria, study design factors and patient characteristics observed between the two studies. For example, brain metastases and previous treatment with targeted therapies for thyroid cancer were exclusion criteria in the DECISION trial but not in the SELECT trial. As these differences might bias the MAIC results, patients fulfilling these criteria were removed from the SELECT population. Other characteristics that were different between the studies were age, sex, Eastern Cooperative Oncology Group performance status, geographical region, histology and site of metastasis. As these characteristics were also potential treatment effect modifiers, the population with IPD (SELECT trial) was adjusted and reweighted to match the population with summary statistical data (DECISION trial), following the approach of Signorovitch *et al.* (2012). Weights were created by performing a logistic regression on the patient-level SELECT data and the summary DECISION data. The resulting predicted values (using propensity score) were used to weight the SELECT data. Complete details on this MAIC process are published elsewhere [30,31].

ORR was assumed to be only treatment related, in other words, not affected by baseline characteristics. The percentages of responders in the placebo arms were close to zero in both trials, suggesting that apart from the treatment effect, no other covariates influence the outcome. ORR is a direct measure of drug antitumor activity. Stable disease, which can reflect the natural history of disease, is not a component of ORR [32]. Therefore, MAIC was not performed on ORR. When correcting for baseline characteristics, the results for PFS and OS were slightly different, but showed the same trend as in the base case analysis (Table 2).

Table 2. Point estimate of number needed-to-treat for lenvatinib versus sorafenib for progression-free survival, overall survival and objective response rate at 24 months before matching adjustment.

Outcome	OR _{L-S}	EER _S	ARR _{L-S} (95% CI)	NNT _{L-S} (95% CI)
PFS				
Delta method	8.594	0.174	-0.470 (-0.716 to -0.224)	-2.13 (-4.47 to -1.40)
Simulation-based	–	–	-0.460 (-0.670 to -0.207)	-2.17 (-4.82 to -1.49)
OS				
Delta method	1.345	0.681	-0.061 (-0.181–0.059)	-16.48 [†] (-∞ to -5.53) ∪ (16.84 ± ∞)
Simulation-based	–	–	-0.058 (-0.172–0.067)	-17.31 [†] (-∞ to -5.81) ∪ (14.95 ± ∞)
ORR				
Delta method	4.332	0.122	-0.254 (-0.827–0.318)	-3.93 [†] (-∞ to -1.21) ∪ (3.14 ± ∞)
Simulation-based	–	–	-0.280 (-0.751–0.078)	-3.57 [†] (-∞ to -1.33) ∪ (12.79 ± ∞)

[†]When the 95% CI of ARR encompasses zero, the 95% CI of NNT is in two pieces.
ARR: Absolute risk reduction; EER: Expected event rate; L-S: Lenvatinib–sorafenib; NNT: Number needed-to-treat; OR: Odd ratio; ORR: Objective response rate; OS: Overall survival; PFS: Progression-free survival; S: Sorafenib.

Table 3. Point estimate of number needed-to-treat for lenvatinib versus sorafenib based on matching adjusted indirect comparison.

Outcome	ACTIVE TX		Placebo		ARR (95% CI)	NNT (95% CI)
	EER	Total	CER	Total		
PFS at 24 months						
SELECT (lenvatinib)	0.443	261	0.043	131	0.400 (0.330–0.470)	2.50 (2.13–3.03)
DECISION (sorafenib)	0.174	207	0.083	210	0.092 (0.028–0.155)	10.93 (6.45–35.91)
OS						
SELECT (lenvatinib)	0.664	261	0.487	131	0.177 (0.074–0.280)	5.64 (3.57–13.46)
DECISION (sorafenib)	0.681	207	0.657	210	0.024 (-0.066–0.114)	41.67 [†] (-∞ to -15.08) ∪ (8.75 ± ∞)

[†]When the 95% CI of ARR encompasses zero, the 95% CI of NNT is in two pieces.
ARR: Absolute risk reduction; CER: Control event rate; EER: Expected event rate; NNT: Number needed-to-treat; OS: Overall survival; PFS: Progression-free survival; TX: Treatment.

Table 4. Point estimate of number needed-to-treat for lenvatinib versus sorafenib based on matching adjusted indirect comparison.

Outcome		OR _{L-S}	EER _S	ARR _{L-S} (95% CI)	NNT _{L-S} (95% CI)
Lenvatinib/sorafenib					
PFS	Delta method	7.556	0.174	-0.440(-0.684 to -0.196)	-2.27 (-5.10 to -1.46)
	Simulation-based	–	–	-0.433 (-0.644 to -0.188)	-2.31 (-5.32 to -1.55)
OS	Delta method	1.868	0.681	-0.118 (-0.223 to -0.014)	-8.44 (-69.29 to -4.49)
	Simulation-based	–	–	-0.116 (-0.215 to -0.007)	-8.62 (-139.96 to -4.65)

ARR: Absolute risk reduction; EER: Expected event rate; L-S: Lenvatinib–sorafenib; NNT: Number needed-to-treat; OR: Odd ratio; OS: Overall survival; PFS: Progression-free survival; S: Sorafenib.

Results

The results for the unadjusted ITC showed that at 24 months, for every two RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have progressed (Table 3). This difference was statistically significant. For OS and ORR, although not statistically significant, the values showed a trend in favor of lenvatinib.

While the point estimates are similar to the unadjusted indirect-NNT, the MAIC results provided statistically significant results for both PFS and OS, in favor of lenvatinib (Table 4). At 24 months, for every two RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have progressed and for every eight RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have died.

Discussion

Reporting results in terms of numbers needed to treat is clinically easy to interpret [13,33,34]. NNT allows the decision makers to estimate the number of patients needed to be treated with one therapy versus another for one

patient to encounter an outcome of interest within a defined period of time. However, NNT is rarely used for ITCs, despite having been referenced as a possible scale in meta-analysis methodology guidelines [17,35]. To our knowledge, this is the first study to use NNT as summary measure for results obtained from an ITC or MAIC analysis. In the case study presented, two active treatments were relevant for treating RR-DTC and the evidence was collected in two individual RCTs, one with individual level data and the other with aggregated data. By simply applying Equation 2, the NNT was estimated. 95% CIs were derived using two novel approaches, one based on the delta method and the other on a simulation-based method.

One limitation of the ITC was the potential violation of the consistency assumption (i.e., transitivity of effect size through a common comparator), as differences in the study designs and patient characteristics (both possible treatment effect modifiers) were identified. Such a limitation motivated the use of MAIC instead. The performance of matching adjustment, which our consistency assumption and consequent derivations are based upon, were integral to the analysis, and have been described elsewhere in detail. The case study presented both unadjusted and adjusted analyses to study the impact of these differences on the NNT results. Despite leveraging all the potential treatment effect modifiers available in the MAIC, the risk that unknown and/or unmeasured characteristics might bias the results remain.

Future studies may extend this work to the indirect comparison of three or more active treatments, or if both direct and indirect evidence are available, to NMA.

One concern about using the NNT is that it is often quoted as a single number without a 95% CI [36]. By applying the delta method or the simulation-based method, we demonstrated that a 95% CI around ARR can be estimated. By inverting the lower and upper limits, this can then serve as the 95% CI around NNTs. However, when the confidence interval of ARR includes zero, the 95% CI around NNT will still consist of a positive piece which goes to infinity and a negative piece which goes to minus infinity [15]. In this case, we suggest quoting the NNT point estimate and its two-piece 95% CI, together with the ARR estimate and its 95% CI.

Conclusion

Including NNT information alongside evidence synthesis findings may provide clarity in the clinical interpretation of the results. The information is presented in an easy-to-comprehend format for the audience. Reporting of a scale that is more clearly interpretable from a clinical perspective may enhance the usefulness and adoption of evidence synthesis.

Summary points

- Evidence synthesis is increasingly used but the odds ratio (OR) summary results are difficult to communicate.
- This article shows how to derive number needed to treat from OR obtained by an indirect treatment comparison or matching adjusted indirect comparison.
- In the illustrative example, the ORs of sorafenib versus lenvatinib progression-free survival and overall survival at 24 months, in radioiodine-refractory differentiated thyroid cancer (RR-DTC) patients, as obtained from matching adjusted indirect comparison were respectively 7.6 and 1.9.
- This means that, for every two RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have progressed and for every eight RR-DTC patients treated with lenvatinib instead of sorafenib, one fewer would have died.

Financial & competing interests disclosure

The work in this paper was undertaken as part of a consultancy work sponsored by Eisai. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

- Merlin T, Tamblyn D, Schubert C. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. Australian Government Department of Health (2016). <https://pbac.pbs.gov.au/>
- Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making (2011). http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/TSD1-Introduction.final_08.05.12.pdf
- Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect evidence: indirect treatment comparisons in meta-analysis. Canadian Agency for Drugs and Technologies in Health, Ottawa, Canada (2009).
- Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *Brit. Med. J.* 321(7256), 255–256 (2000).
- Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat. Med.* 22, 2995–3016 (2003).
- Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Brit. Med. J.* 331, 897–900 (2005).
- Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat. Med.* 23, 3105–3124 (2004).
- Lu G, Ades A. Assessing evidence consistency in mixed treatment comparisons. *J. Am. Stat. Assoc.* 10, 447–459 (2006).
- Signorovitch JE, Wu EQ, Yu AP et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparison applied to psoriasis treatment with adalimumab or etanercept. *Pharmacoeconomics* 28(10), 935–945 (2010).
- Signorovitch JE, Sikirica V, Erder MH et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 15(6), 940–947 (2012).
- Hutton B, Salanti G, Caldwell DM. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann. Intern. Med.* 162, 777–784 (2015).
- Cochrane Handbook for Systematic Reviews of Interventions*. Higgins JPT, Green S (Eds). The Cochrane Collaboration, London, UK (2008).
- Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *Brit. Med. J.* 310(6977), 452–454 (1995).
- Faraone SV. Interpreting estimates of treatment effects implications for managed care. *Pharm. Ther.* 33(12), 700–711 (2008).
- Altman DG. Confidence intervals for the number needed to treat. *Brit. Med. J.* 317(7168), 1309 (1998).
- Schechtman E. Odds ratio, relative risk, absolute risk reduction, and the number needed to treat – which of these should we use? *Value Health* 5, 431–436 (2002).
- Dias S, Sutton AJ, Ades AE. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med. Decis. Making* 33(5), 607–617 (2013).
- Schlumberger M, Tahara M, Wirth LJ et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *N. Engl. J. Med.* 372, 621–630 (2015).
- CLINICAL STUDY REPORT of study of (E7080) lenvatinib in differentiated cancer of the thyroid a multicenter, randomized, double-blind, placebo-controlled, Phase III trial of lenvatinib (E7080) in I-refractory differentiated thyroid cancer. Eisai, April 2014 (Unpublished).
- Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat. Med.* 19(13), 1707–1728 (2000).
- Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat. Med.* 21, 1575–1600 (2002).
- Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence Based Med.* 1(6), 164–166 (1996).
- Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the ‘number needed to treat’? An empirical study of summary effect measures in meta-analyses. *Int. J. Epidemiol.* 31(1), 72–76 (2002).
- Bender R. Calculating confidence intervals for the number needed to treat. *Control. Clin. Trials* 22(2), 102–110 (2001).
- Casella G, Berger RL. *Statistical Inference*. Duxbury, CA, USA (2002).
- Hildebrandt M, Bender R, Gehrman U. Calculating confidence intervals for impact numbers. *BMC Med. Res. Methodol.* 6, 32 (2002).
- Altman DG. Confidence intervals for the number needed to treat. *BMJ* 317(7168), 1309–1312 (1998).
- Bender R. Calculating confidence intervals for the number needed to treat. *Controlled clinical trials* 22(2), 102–110 (2001).
- Brose MS, Nutting CM, Jarzab B et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, Phase III trial. *Lancet* 384(9940), 319–328 (2014).
- Tremblay G, Li X, Abouzaid S et al. Number needed to treat (nnt) analysis of therapies in radioiodine refractory differentiated thyroid cancer (rr-dtc) using indirect comparison. Poster presented at: *The 15th International Thyroid Congress (ITC)*. Lake Buena Vista, FL, USA, 18–23 October (2015).

31. Tremblay G, Holbrook T, Milligan G. Matching-adjusted indirect treatment comparison in patients with radioiodine-refractory differentiated thyroid cancer. *Comp. Eff. Res.* 6, 13–21 (2016).
32. Guidance for industry: clinical trial end points for the approval of cancer drugs and biologic. FDA (2007). www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf
33. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: absolute risk reduction, relative risk reduction, and number needed to treat. *Persp. Clin. Res.* 7(1), 51–53 (2016).
34. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *Pharm. Ther.* 33(12), 700–711 (2008).
35. Jansen J, Fleurence R, Devine B *et al.* Interpreting indirect treatment comparisons and network meta-analysis for healthcare decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value Health* 14, 417–428 (2011).
36. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann. Intern. Med.* 126, 712–720 (1997).