Methodologies for network meta-analysis of randomised controlled trials in pain, anaesthesia, and perioperative medicine: a narrative review

Brett Doleman[1,*] , Janus Christian Jakobsen[2], Ole Mathiesen[3,4] , Nicola Cooper[5],
Alex Sutton[5] and Jonathan Hardman[1]

[1]University of Nottingham, Nottingham, UK, [2]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Capital Region of Denmark & Department of Regional Health Research, The Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark, [3]Centre for Anaesthesiological Research, Department of Anaesthesiology, Zealand University Hospital Køge, Køge, Denmark, [4]Department of Clinical Medicine, Copenhagen University, Copenhagen, Denmark and [5]Complex Reviews Synthesis Unit (CRSU), Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK

*Corresponding author. E-mail: dr.doleman@gmail.com

Summary

Network meta-analysis has emerged as a method for analysing clinical trials, with a large increase in the number of publications over the past decade. Network meta-analysis offers advantages over traditional pairwise meta-analysis, including increased power, the ability to compare treatments not compared in the original trials, and the ability to rank treatments. However, network meta-analyses are inherently more complex than pairwise meta-analyses, requiring additional statistical expertise and assumptions. Many factors can affect the certainty of evidence from pairwise meta-analysis and can often lead to unreliable results. Network meta-analysis is prone to all these issues, although it has the additional assumption of transitivity. Here we review network meta-analyses, problems with their conduct and reporting, and methodological strategies that can be used by those conducting reviews to help improve the reliability of their findings. We provide evidence that violation of the assumption of transitivity is relatively common and inadequately considered in published network meta-analyses. We explain key concepts with clinically relevant examples for those unfamiliar with network meta-analysis to facilitate their appraisal and application of their results to clinical practice.

Keywords: network meta-analysis; review methodology; statistics in anaesthesia; systematic review; transitivity

---

Editor's key points

- Network meta-analyses (NMAs) provide a synthesis of evidence that can be used to inform treatment guidelines, so it is important to ensure that they are conducted and interpreted with the highest standards.
- Network meta-analysis offers many advantages over pairwise meta-analysis, but is inherently more complex, with the important additional assumption of transitivity.

- The authors review evidence from published studies of violations of this assumption, coupled with inadequate conduct of NMAs in the field of anaesthesia and other specialties.
- Approaches to improve the conduct and interpretation of NMAs are critical for the preparation, appraisal, and clinical application of NMAs in pain, anaesthesia, and perioperative medicine.

The number of published network meta-analyses (NMAs) has greatly increased in recent years, increasing by a factor of 20 from 2010 to 2017.[1] Network meta-analyses can be used to inform treatment guidelines[2–4]; it is therefore of utmost importance to ensure that their conduct and reporting are of the highest standards.[5] Network meta-analysis offers advantages over pairwise meta-analysis (PWMA), including increased power, the ability to compare treatments not compared in the original trials, and the ability to rank treatments.[6] This review discusses the methodological considerations specific to NMA.

Issues that affect the certainty of evidence in PWMA include risks of bias, random errors, publication bias, and heterogeneity in the included studies.[6] NMAs can also suffer from all these limitations, which can manifest as poor ability to predict results from large randomised controlled trials (RCTs)[7] and a scarcity of high-quality evidence.[8] However, NMA has an additional fundamental assumption known as transitivity, which states that study characteristics that influence the effectiveness of the interventions are balanced between different treatment comparisons. Here we use simple clinical examples from acute pain studies to illustrate principles of NMA with simplified equations aimed at clinicians not familiar with statistics. After discussing what an NMA is and aspects of the methodology that should be considered to reduce erroneous conclusions, we present evidence from meta-epidemiological studies to support arguments for the limitations in the conduct of NMAs, both specific to

anaesthesia and for other specialties. We aim to help authors of NMAs improve their methodology and help clinicians appraise such articles to guide their practise.

What is network meta-analysis?

Network meta-analysis is a meta-analytic technique that estimates treatment effects from direct ($_{DIR}$) and indirect ($_{IND}$) estimates.[9] Direct effects are obtained when two interventions have been compared within a trial (trial one compares A *vs* B and trial two compares A *vs* C). Indirect effects can then be calculated, even when interventions are not directly compared within a trial (C *vs* B), by computing the difference between each intervention (B *vs* C) and a third intervention, called a common comparator (A).

As a specific example, we imagine an NMA of nonopioid analgesics for treatment of acute pain that includes four infinitely large trials that use 24-h morphine consumption as their outcome. Although pain might be a more clinically relevant outcome,[10] we use morphine consumption as our previous data suggest it has the appropriate statistical characteristics to help illustrate our arguments. Two trials include placebo *vs* gabapentin, whilst the other two trials include placebo *vs* lidocaine (Fig. 1a). Using a PWMA, we can only include mean difference (MD) estimates from trials where interventions are directly compared. However, in an NMA we can compare the effects of both active treatments by looking at the difference between each active drug and placebo. This is given by
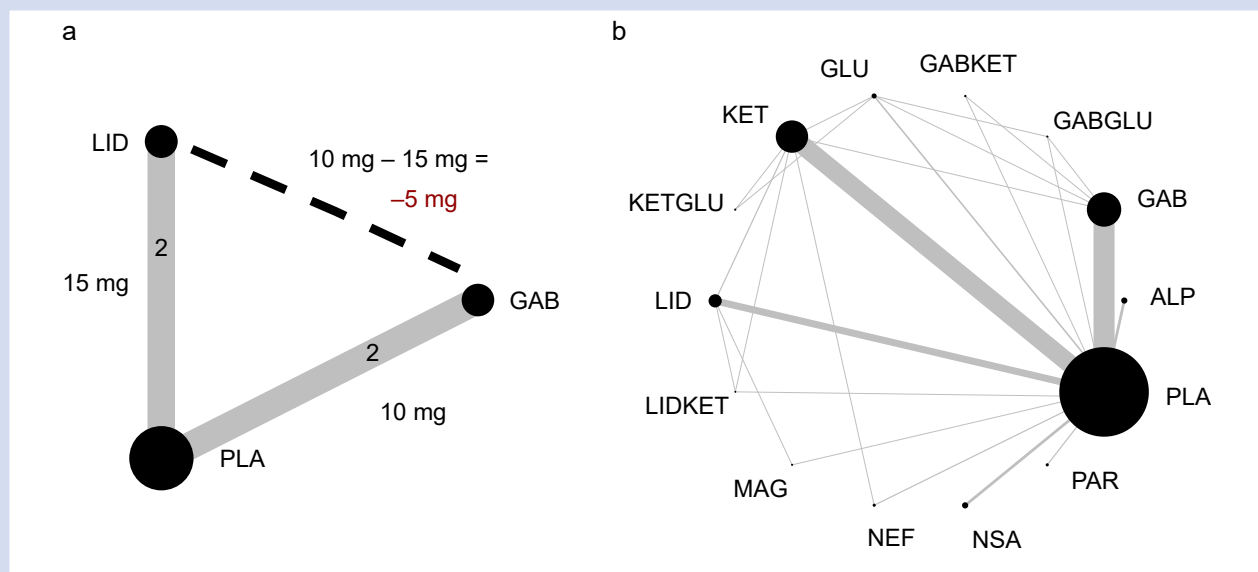
Fig 1. (a) Network plot showing the hypothetical network meta-analysis. Each direct comparison of lidocaine (LID) and gabapentinoids (GAB) with placebo (PLA) is shown. The dashed line shows the indirect comparison between GAB and LID. Using the formulae in the text, an estimated reduction of 5 mg in morphine dose with LID can be achieved based on indirect comparisons derived from direct comparisons with a common comparator. (b) Network plot of interventions for reducing the incidence of chronic postsurgical pain (CPSP).[12] Compared with (a), there are loops of evidence showing direct and indirect estimates for GAB, ketamine (KET), and PLA. This can be used to assess inconsistency as opposed to the evidence from the same plot for previous interventions as there is no trial comparing these agents directly. ALP, alpha-2 agonists; GAB, gabapentinoids; GABGLU, gabapentinoids and glucocorticoids; GABKET, gabapentinoids and ketamine; GLU, glucocorticoids; KET, ketamine; KETGLU, ketamine and glucocorticoids; LID, lidocaine; LIDKET, lidocaine and ketamine; MAG, magnesium; NEF, nefopam; NSA, NSAIDs and COX-2 inhibitors; PAR, paracetamol; PLA, placebo.

equation (1). We have excluded sampling error and the possibility of heterogeneity[11] to help with interpretation of the equations by including infinitely large trials:

$$MD_{IND\ lidocaine\ vs\ gabapentin} = MD_{DIR\ placebo\ vs\ gabapentin}$$
$$- MD_{DIR\ placebo\ vs\ lidocaine} \quad (1)$$

In this example, we will assume that the effects of gabapentin in two trials are a reduction in morphine consumption of 10 mg compared with placebo; for lidocaine, the reduction compared with placebo is 15 mg. We can estimate the indirect effect of lidocaine vs gabapentin as being a reduction of 5 mg in favour of lidocaine using the above formula:

$$MD_{IND\ lidocaine\ vs\ gabapentin} = MD_{DIR\ placebo\ vs\ gabapentin}$$
$$- MD_{DIR\ placebo\ vs\ lidocaine} = 10\ mg - 15\ mg = -5\ mg \quad (2)$$

This process can be repeated multiple times in more complex networks to yield multiple indirect estimates via multiple common comparators (Fig. 1b). After this calculation of indirect estimates, both direct (if available) and indirect estimates can be pooled together to give a mixed estimate. By combining two estimates rather than one, statistical power and clinical reliability are potentially increased. Although this can be thought of conceptually and shown mathematically,[13] this increase in power has also been formally demonstrated empirically.[14,15]

We briefly discuss the calculation of variance (sampling error) to demonstrate the contribution made by indirect estimates to mixed estimates. For indirect estimates, variance (VAR) is calculated by adding the variances of each of the direct estimates above (2)[15] This means that the indirect estimate variance will always be higher than the direct estimates they are based on. To illustrate, we assume that the variances from both direct estimates are equal to one; indirect variance is described by:

$$VAR_{IND\ lidocaine\ vs\ gabapentin} = VAR_{DIR\ placebo\ vs\ gabapentin}$$
$$+ VAR_{DIR\ placebo\ vs\ lidocaine} = 1 + 1 = 2 \quad (3)$$

This variance can increase to large values as the number of estimates increases. Alternatively, if we calculate the variance for two studies directly comparing lidocaine and gabapentin using the same variance as the previous example, the variance is lower:

$$VAR_{DIR\ lidocaine\ vs\ gabapentin} = (1/1 + 1/1)^{-1} = 0.5 \quad (4)$$

Because of the addition of indirect effects, NMA has advantages over traditional PWMAs, which can only include direct estimates. Firstly, although we have discussed the higher variance of indirect estimates when compared with direct estimates, the addition of indirect estimates contributes additional information so statistical power is improved.[14,15] Secondly, NMA allows comparison of agents not compared within the original trials. Although large, well-conducted RCTs are the gold standard for assessing interventions, a large RCT comparing all treatments might be unfeasible and prohibitively expensive, so NMA offers an alternative approach. An example of this would be an NMA examining antidepressant therapy, which included >20 interventions and >100 000 participants; this would be an extremely challenging goal for an RCT.[16] Thirdly, NMA allows use of statistical techniques that can help identify potential best treatments, such as

ondansetron and dexamethasone combinations after intrathecal morphine.[17] NMA can also include multi-arm trials whilst reducing the risk of double counting in multiple pairwise comparisons.[18] Finally, NMA can consider the entire evidence base in a systematic fashion and allows formal investigation of inconsistency. For example, if results from multiple PWMAs are compared informally, transitivity can still be an issue (discussed below), but NMA allows this to be investigated and quantified formally, aiding clinical decision-making.

Disadvantages of NMAs include their inherent statistical complexity; it is therefore recommended that NMAs be conducted with a statistician experienced in NMA.[19] There are also concerns over NMAs being largely informed by indirect evidence. A previous analysis of 213 NMAs found that indirect evidence was responsible for the majority of the information used to inform treatment effects.[20] There is debate over whether direct evidence should be preferred to indirect estimates. Previous studies have identified problems with low power of indirect estimates.[21] If indirect effects are calculated from unbiased direct estimates, indirect estimates should also be unbiased (assuming transitivity).[22]

However, indirect estimates are inherently observational because individual trials are not randomised to treatment options, in contrast with the trial participants. Therefore, as with observational studies, confounding can be an issue (see below). However, it should also be noted that indirect estimates usually contribute less information if comparably sized direct evidence is available because of how variance is calculated. Detailed analyses of this issue have concluded that they are generally equally susceptible to bias except in particular circumstances.[23]

Systematic review methods

As with PWMA, a well-defined research question is required in population, intervention, comparison, outcome, and statistical analysis (PICOS) format. In particular, included interventions should be defined *a priori,* and the authors should make explicit which interventions will be grouped and treated as identical, which we call a treatment node. Placebo trials should be included as they usually provide the largest number of studies and are a good source of indirect evidence. Indeed, exclusion of these studies can have large effects on NMA effect estimates.[24] Also, older interventions should be considered even if they are used less frequently in clinical practice as they can provide valuable indirect evidence.[25]

Careful consideration must be taken when grouping interventions, which needs a clinical rationale. The more diverse the interventions included in a node, the greater the risk of clinical heterogeneity. However, this is offset by an increase in precision in estimates and network connectedness. The type of study for inclusion should also be considered. RCTs remain the gold standard for assessing interventions. If comparative observational studies are to be considered, specific methods have been proposed for incorporation of non-randomised evidence.[26] Recent evidence shows that comparative observational evidence can produce similar estimates to RCTs,[27] although this practice remains controversial.

The systematic review portion of the NMA should be conducted thoroughly by searching multiple sources and involving a clinical librarian to improve retrieval of relevant studies.[28] This was performed in only 19% of NMAs in

anaesthesia.[9] Concerns have been raised regarding exclusion of non-English language articles causing English language bias. However, recent research concluded that this makes little difference to review effect estimates.[29] Nevertheless, the finding that the statistical significance of results changed 9% of the time should be considered to represent an impactful influence, and therefore non-English language studies should be included.[29]

An extension of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist called PRISMA-NMA[30] should be used to improve the quality of reporting. However, a previous study that included more than 1000 NMAs found little effect of its introduction on the quality of reporting in NMAs over and above previous yearly increases.[31] Poor reporting of the checklist is more common for industry-funded NMAs.[32] Therefore, journals and peer reviewers must ensure this has been adequately completed to improve reporting, particularly as only 42% of NMAs in anaesthesia used the PRISMA-NMA checklist.[9]

As with traditional PWMA, registration of a protocol with statistical analysis, either for publication or on a database such as PROSPERO, is essential, and this is associated with higher review quality.[33] However, <50% of NMAs in anaesthesia have a pre-registered NMA protocol.[9] Because of the complex nature of the analysis, involvement of an experienced statistician should be implemented, although this was found in only 19% of NMAs in anaesthesia.[9] Statistician involvement at the protocol stage will enable development of a robust statistical analysis plan.

Quality of network meta-analyses

As with traditional PWMA, tools are available for authors to ensure that reported estimates can be presented with levels of certainty and confidence based on deficiencies in domains, as discussed in this article. An extension of Grading of Recommendations Assessment, Development and Evaluation (GRADE) is available, which incorporates issues specific to NMA.[34] An alternative is Confidence in Network Meta-Analysis (CINeMA); this uses online software that allows data to be uploaded and automates many of the assessments required.[35] GRADE and CINeMA can produce different ratings around a third of the time, with CINeMA being more conservative.[36] A third option is a type of threshold analysis.[37]

Historically, reporting in NMA has been inadequate.[38] Although the quality of NMAs has increased over the years, significant limitations still exist.[39] A review of 438 NMAs found only a quarter were of high quality, and a fifth were of poor quality.[40] A review of 62 anaesthetic NMAs[9] found deficiencies in the areas of study searches, team composition, reporting of methodology, and reporting of many types of statistical analysis. Overall, 21 out of 29 reporting domains had low or very low compliance. Using the AMSTAR-2 tool, 84% of anaesthesia NMAs were rated as critically low, with quality scores increasing with journal impact factor and use of PRISMA-NMA.[41]

Transitivity and incoherence (inconsistency)

Transitivity is the major additional assumption in NMA. It assumes that effect modifiers are equally distributed between different treatment comparisons. Alternatively, it can be defined as any eligible participant should be jointly randomisable to any of the included interventions. If this assumption is violated, avoidance of NMA or use of strategies discussed later in this section should be considered.[22]

To illustrate this assumption using the previous example (where we assumed effect modifiers were balanced), assume that new trials of gabapentin *vs* placebo now include participants at high risk of postoperative pain (females with pre-existing pain) and all the lidocaine *vs* placebo trials include participants at low risk of pain (males undergoing laparoscopy). Because previous data have illustrated that reductions in morphine consumption are determined by the underlying risk of pain (Fig. 2a),[42–44] it would not be valid to compare these agents indirectly under these circumstances.[45] This is because efficacy is now determined both by the difference in efficacy between the two drugs and the difference in the distribution of this effect modifier (Fig. 2b). To demonstrate this mathematically, we assume the difference in this effect modifier causes a 5-mg reduction in morphine consumption benefitting gabapentin, which we input into equation (2):

$$\mathrm{MD_{IND\ lidocaine\ \textit{vs}\ gabapentin}} = \mathrm{MD_{DIR\ placebo\ \textit{vs}\ gabapentin}}$$
$$- \mathrm{MD_{DIR\ placebo\ \textit{vs}\ lidocaine}}$$
$$+ \Delta\mathrm{effect\ modifier_{lidocaine\ \textit{vs}\ gabapentin}}$$
$$= 10\ \mathrm{mg} - 15\ \mathrm{mg} + 5\ \mathrm{mg} = 0\ \mathrm{mg} \tag{5}$$

This results in no difference in efficacy in our hypothetical example where we know (unlike real life) that lidocaine results in a 5-mg reduction compared with gabapentin (2). This is termed confounding in observational evidence but also extends to NMA. A previous study has identified that re-analysing NMAs using covariate adjustment (i.e. underlying risk of pain in this context) can alter the results.[46] Therefore, adjustment for covariates remains a practice that can mitigate inconsistency, although it can have low power if the number of included trials is low.[47] Despite this low power, covariate adjustment should still be used because of the issue above.

Many methods have been proposed to test the transitivity assumption, each with advantages and disadvantages.[48] However, these methods rely on knowledge of potential effect modifiers when testing the assumption and cannot account for unknown effect modifiers ('unknown unknowns'). Before analysis, graphical displays can be used to illustrate the distribution of known effect modifiers between different treatment comparisons, similar to the contribution matrices used in CINeMA.[49] Potential effect modifiers can also be identified using a network meta-regression using trial-level covariates. If it is found that the transitivity assumption has been violated, options include not performing an NMA or running a network meta-regression with the specific effect modifier and presenting estimates from a fixed value of this covariate.[50] In our example, baseline risk of pain can be estimated from control group values to prevent variations violating transitivity.[12]

Although related, incoherence (inconsistency) describes statistically significant differences between direct and indirect treatment estimates. Evaluation of this requires closed loops of evidence (Fig. 1b). Using our previous example, if two RCTs were conducted comparing gabapentin and lidocaine directly and these showed a 5-mg reduction with lidocaine compared with gabapentin, then:

$$\mathrm{MD_{IND\ lidocaine\ \textit{vs}\ gabapentin}} = \mathrm{MD_{DIR\ lidocaine\ \textit{vs}\ gabapentin}}$$
$$= -5\ \mathrm{mg} \tag{6}$$
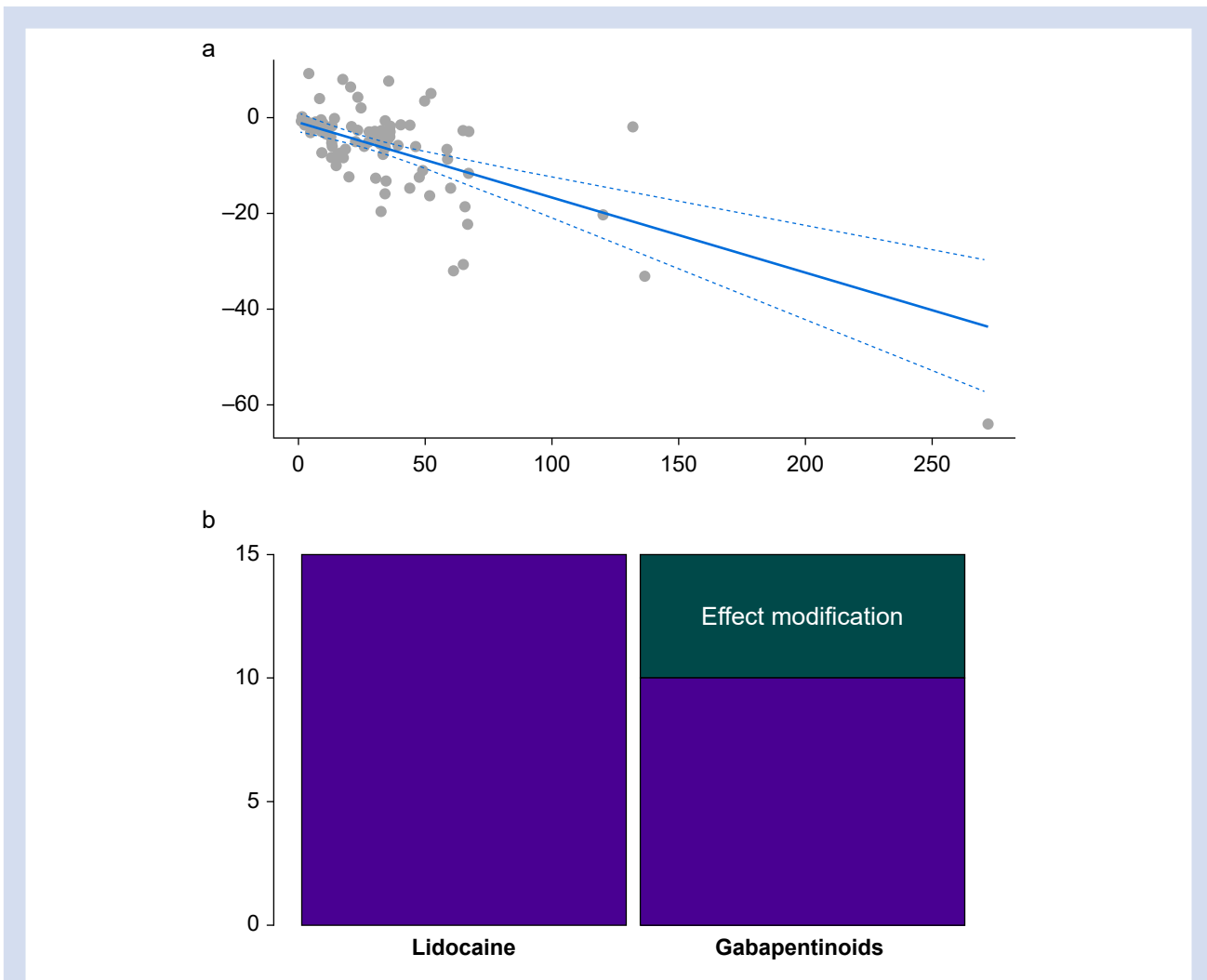
Fig 2. (a) Network meta-regression of α_2 adrenergic agonists for the outcome of 24-h morphine consumption. The *x*-axis is baseline risk (control group morphine consumption), and the *y*-axis is the mean reduction in 24-h morphine consumption. As baseline risk increases, efficacy also increases. Grey dots are the effects for each study, the solid blue line is the regression line estimate, and the dashed blue line is the 95% confidence interval (CI). Clinical significance changes with the value of baseline risk (*x*), so a single conclusion cannot be drawn on clinical significance. Baseline risk is a potential effect modifier, and unequal distribution between comparisons could violate transitivity, as shown. (b) Stacked bar chart showing the absolute morphine consumption after gabapentinoids (GAB) and lidocaine (LID) relative to placebo. The purple regions highlight the true difference in morphine consumption between the GAB and LID groups (i.e. LID more effective). The green region shows the effect of a difference in the distribution of effect modifiers (baseline risk or risk of bias) between comparisons of active interventions with placebo. Despite the 'known' underlying benefit of LID (purple region), a difference in effect modifiers (green region) makes them falsely appear to have the same efficacy.

Here, our indirect and direct estimates match, and consistency can be reasonably assumed. Alternatively, if the same two RCTs showed a 10-mg reduction with gabapentin, then:

$$\text{MD}_{\text{IND lidocaine } vs \text{ gabapentin}} = -5 \text{ mg}$$
$$\neq \text{MD}_{\text{DIR lidocaine } vs \text{ gabapentin}} = 10 \text{ mg} \tag{7}$$

This suggests incoherence. In practice, estimates are not from infinitely large trials and are unlikely to match exactly, so inferential statistics are performed, and consistency is assumed if a certain threshold of statistical significance is

observed. However, lack of statistical incoherence does not necessarily indicate satisfaction with the transitivity assumption because of the low power of these tests.

In general, tests for incoherence can be classified as local or global. Local tests assess local loops of evidence, whereas global tests assess the network as a whole. Examples of local tests include measurements of the absolute differences between direct and indirect estimates.[51] Direct and indirect estimates should be presented together in a forest plot to allow visual inspection for incoherence as statistical tests have low power.[52] However, the power of these tests improves with increasing sample sizes and lower heterogeneity in the

included studies.[53] Another visual example includes net heat plots, although their reliability has been questioned.[54] An example of a global test for incoherence is the design-by-treatment interaction.[55] Review authors should use both approaches to inform concerns over incoherence.[19]

Because it is such an essential assumption of NMA, it is useful to estimate the incidence of incoherence in the NMA literature. A study from 2013 found an incidence of local incoherence of 2–9%.[56] More recently, a study re-analysed 201 NMAs and found global inconsistency in up to 20% of reviews.[57] Another older analysis of 112 trial networks reported an incidence of inconsistency of 14%.[58] It is also useful to evaluate how often NMA authors are assessing for transitivity and incoherence. A recent analysis of 721 NMAs found that reviews were more than three times as likely to assess transitivity statistically rather than conceptually.[59] Even within Cochrane reviews, inconsistency is only assessed around a third of the time.[60] Another analysis of 77 NMAs found that almost half did not consider effect modification in their analyses.[61] Similarly, around half of 438 NMAs did not assess or account for inconsistency between direct and indirect treatment effects.[40] More specific to NMAs in anaesthesia, a recent analysis showed that transitivity was only tested in 65% of NMAs.[9]

Risk of bias

Although we have discussed risk of bias (ROB) in underlying RCTs for PWMAs,[6] a further consideration in NMA is the contribution that trial data make to each treatment comparison. For example, using our previous example, if all trials comparing gabapentin *vs* control are unblinded and all trials comparing lidocaine with placebo have a low ROB, this might exaggerate the effect of gabapentin, which could threaten the transitivity assumption in the same way as baseline risk. CINeMA includes online software that allows the user to upload their data to yield a contribution matrix that can identify ROB in different treatment comparisons.[35] Because of the effects that ROB can have on treatment effects, we recommend that NMAs include only analyses with low-ROB trials, even for a sensitivity analysis.

Heterogeneity

As with PWMA, the clinical heterogeneity of included trials should be assessed, and NMA should only be conducted if trials are sufficiently similar. It is unlikely that many RCTs from different countries and centres will have similar patient populations, which can introduce a large amount of clinical heterogeneity. This needs to be carefully considered, and if NMA is performed, random-effects models are most likely to satisfy the assumptions of included studies in a typical NMA.[6] Clinical heterogeneity requires careful decisions on whether treatments are sufficiently similar to be included in the same group (node); this relates to, for example, different drugs with the same mechanism of action or different doses of the same drug (which can be explored using a model-based NMA[62]). This issue highlights the importance of having clinical experts as part of the review team. If combination treatments (gabapentin + lidocaine) are included, component NMA can allow estimation of the effects of combinations of interventions, even if these were not used in the original trials.[63]

Probably the most intuitive method to present statistical heterogeneity is 95% prediction intervals (PIs), which describe a range within which future treatment effects should be expected and can be used in NMA.[64] Use of 95% PIs provides precision despite sampling error and statistical heterogeneity; they are, therefore, usually wider than the corresponding intervals based on sampling error only (confidence intervals [CI] or credible intervals [CrIs]). Usefully, these are calculated and applied automatically as part of CINeMA online software when assessing the certainty of evidence.[35] Incoherence and heterogeneity are similar concepts; they both describe more variability than expected by chance between effect sizes. The difference is that heterogeneity describes this within the same comparison, and incoherence describes this between comparisons.

Publication bias

Publication bias refers to the preferential publication of positive studies, which has the potential to bias the results of an NMA. Meta-epidemiological studies have identified the highly prevalent nature of publication bias within the anaesthesia literature, with an incidence as high as 43–80%.[65] A study of the proportion of anaesthesia conference abstracts proceeding to full publication found that abstracts with positive findings were significantly more likely to proceed to publication ($P<0.001$).[66] Mandatory trial registration on clinical trial databases can help improve the number of unpublished anaesthesia abstracts that proceed to publication[67] and potentially reduce publication bias.[68] However, 92% of registered RCTs in anaesthesia had an outcome discrepancy between protocols and publications (selective outcome reporting).[69]

The most important factor to reduce the risk of publication bias is a thorough search of unpublished sources.[6] However, only half of NMAs in anaesthesia searched clinical trial databases, and only a fifth searched conference abstracts.[9] This is also a major issue in PWMA in anaesthesia.[70] A study from 2017 in the general literature found that only 36% of 438 NMAs assessed for possible publication bias, and only 40% of NMAs in anaesthesia assessed for publication bias via the treatment network.[9]

Methods for assessing possible publication bias in NMAs include selection models,[71] comparison-adjusted funnel plots, and the approach proposed by Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN).[72] Visual inspection of funnel plots can be inaccurate and should therefore be supplemented with quantitative tests.[73] Comparison-adjusted funnel plots (Fig. 3) are a modification of traditional funnel plots used in PWMA. However, they require a hypothesis on how publication bias manifests, such as novel *vs* old treatments. Therefore, the comparisons to be included in the plot need to be specified.

Imprecision and random error

Power calculations are available for NMAs.[74] Concerns over power (type II errors) and multiplicity (type I errors) occur with NMAs. This is not only because of the potential for a new analysis with the publication of a subsequent RCT, but also because of the multiple comparisons and therefore statistical tests being performed within a single NMA. A previous simulation study found that when simulating 1000 NMAs under the assumption that all treatments were equally effective, >70% of the time at least one treatment comparison demonstrated superiority.[75]

Review authors should carefully interpret effect estimates and the associated measure of imprecision. Null hypothesis statistical testing with *P*-value thresholds unfortunately leads to all-or-nothing statements on the likely effect of an
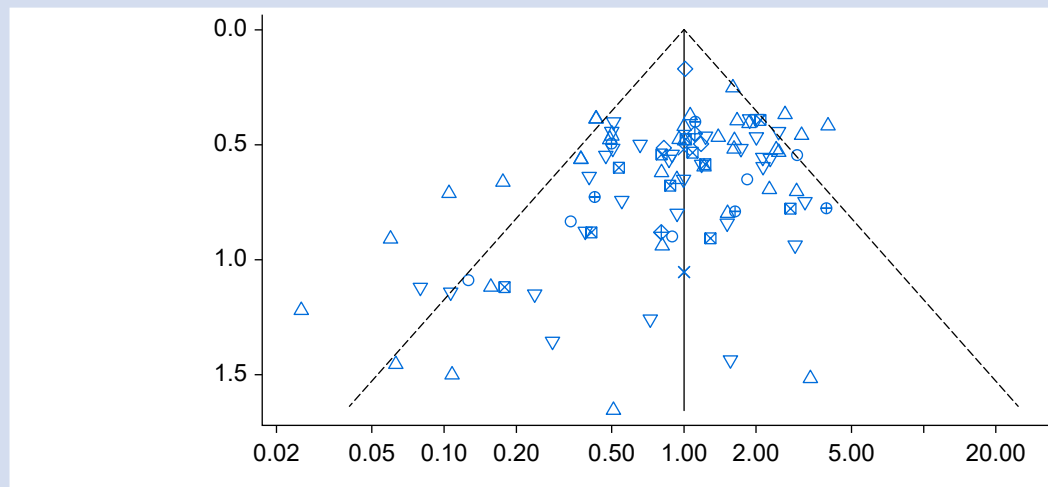
Fig 3. Comparison adjusted funnel plot for nonopioid analgesics in reducing the incidence of chronic postsurgical pain.[12] The y-axis represents the standard error on a reverse scale, and the x-axis represents the effect in each study, a direct summary effect for that comparison. The specific comparison here is active *vs* control interventions, with a different symbol for each comparison (blue points). The funnel plot appears asymmetric, which is corroborated by the test for funnel plot asymmetry (P<0.001).

intervention. Although CIs have been proposed as a preferred method,[76] they are often used similarly to P-values. As an example, we use two identical effect estimates with different magnitudes of precision to highlight this important issue. For an outcome such as postoperative nausea and vomiting (PONV) with an identical intervention, one NMA has a relative estimate of 0.83 (95% CI 0.15−4.64), and another has an estimate of 0.83 (95% CI 0.79−0.88). Despite an identical estimate, the second estimate is more precise, which can be regarded as a reduction in PONV. The first estimate does not necessarily indicate 'no difference', rather that the estimate is too imprecise to draw any conclusions. A recent, less extreme example relates to estimates of the impact of benzodiazepines on delirium.[77] Such assessments are usefully incorporated into CINeMA confidence estimates.

Presentation of network geometry

The geometry of an NMA should be presented with a network plot and, ideally, supplemented with a table of network characteristics not conveyed in the plot. This can give information on the number of studies and participants, and a visual representation of the amount of direct and indirect comparisons included. A well-connected network is ideal as it provides both direct and indirect evidence (Fig. 1b), which helps in the assessment of inconsistency and can also help dilute biased estimates within a network.[78] In addition to demonstrating network geometry,[79] network plots can be modified to demonstrate the distribution of important effect modifiers in the evaluation of the transitivity assumption.[80,81] A contribution matrix[82] can also be used to identify the amount of direct evidence that contributes to each comparison.

Presentation of effect estimates and ranking of treatments

Because an NMA can generate a number of effect estimates, presentation of these can be complicated. Because the general readership is likely to have more experience and familiarity with forest plots in PWMA, a comprehensible summary of NMA results can involve presentation of each intervention in a summary forest plot in comparison with a standard treatment node (placebo). League tables can be used to provide a matrix of effect estimates comparing all treatment combinations. In larger NMAs, these can become complicated and might be more suited to supplementary data. Variations on these can include plots for multiple outcomes,[83] which can help balance interventions with both beneficial and harmful outcomes, forest plots displaying additional information,[84] or plots displaying clinically significant thresholds.[85]

NMA allows the ranking of different treatments. Methods to undertake this include mean/median ranks and cumulative ranking probabilities such as surface under the cumulative ranking curve (SUCRA). The SUCRA value sums all the cumulative probabilities and gives a probability for whether a treatment is among the best options (Fig. 4a). Although initially performed using Bayesian methods, it can be calculated within a frequentist framework using resampling procedures. A frequentist alternative is the P-score that gives similar results to SUCRA without resampling.[86] These can estimate the probability that a treatment is ranked the best. Rankograms can provide a visual representation of these ranking probabilities and give a probability that a treatment is ranked at a certain position (Fig. 4b).

Although the ability to rank treatments is one of the main advantages of NMAs, such results can cause confusion and are often not presented with imprecision. For example, a previous study of 58 NMAs found that by using SUCRA, the median width of 95% CrIs was 65%, and in a third of NMAs there was a >50% probability that the best-ranked treatment was not, in fact, the best.[87] It is therefore suggested that treatment rankings should be presented in conjunction with treatment effects and the uncertainty around both estimates.[88,89]
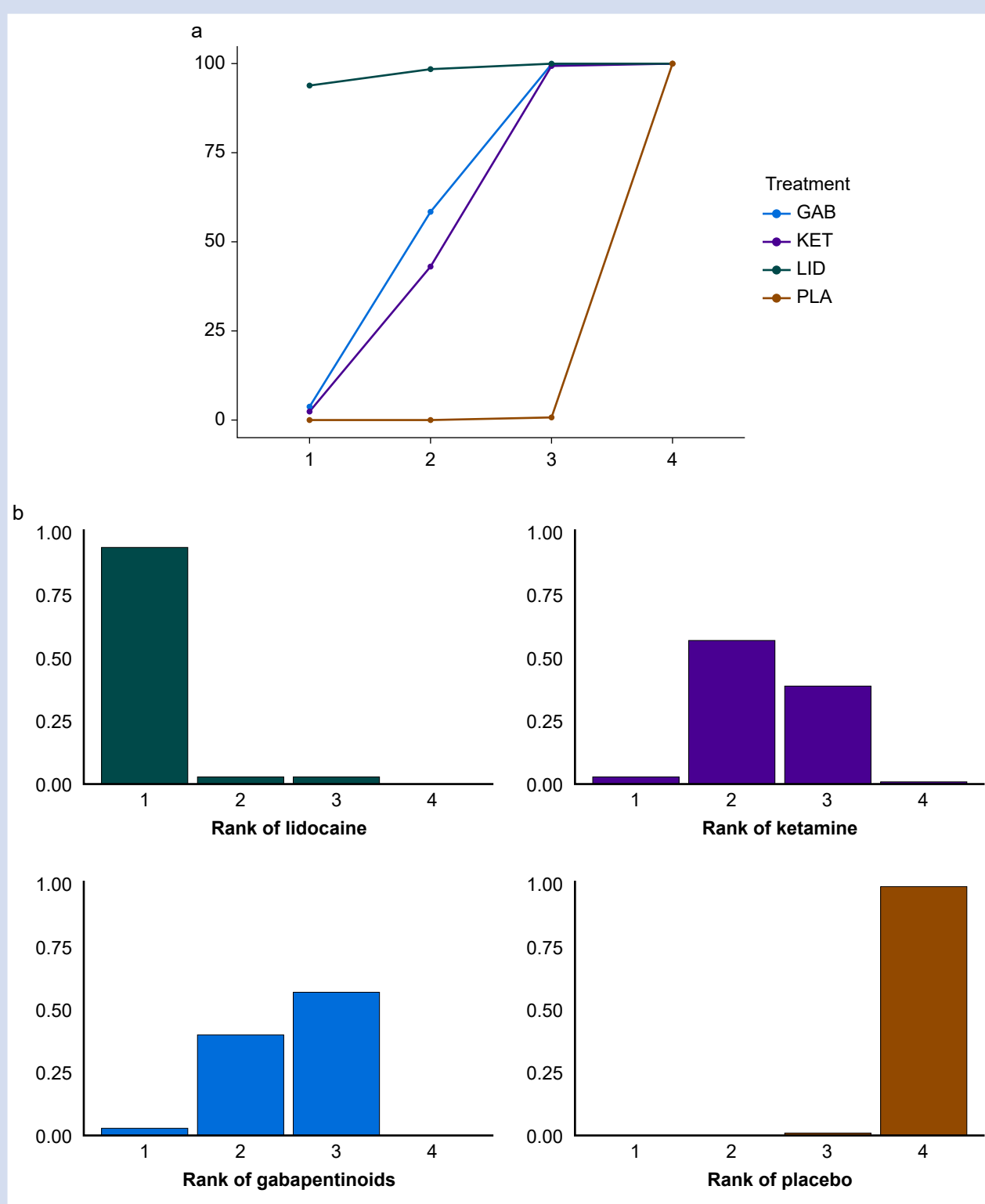
Fig 4. (a) Surface under the cumulative ranking curve (SUCRA) plot for the reduction in chronic postsurgical pain for lidocaine (LID), ketamine (KET), and gabapentinoids (GAB).[12] The area under the curve normalised from 0 to 1 gives the probability that the treatment is ranked as the best (or one of the best). Lidocaine has the largest area (close to 1) and therefore is likely to be the most effective in reducing chronic postsurgical pain (CPSP). Placebo (PLA) has the lowest area under the curve and is therefore likely to be the least effective. (b) Rankogram including four nonopioids for reducing CPSP (using frequentist SUCRA simulations) showing the probability of each treatment being at a certain rank. Lidocaine is ranked first with a probability of 0.94. Ketamine is ranked at least second, with a probability of 0.63 and with a corresponding probability of 0.46 for gabapentinoids. Placebo has a probability of 1 (i.e. certain) of being ranked last.

Bayesian *vs* frequentist analysis

There are a range of software programmes available to assist in the performance of NMA. Packages in R can perform both Bayesian (gemtc[90] and BUGSnet[91]) and frequentist analysis (netmeta[92]). There are also web packages that make analysis easier, without the need for coding skills (MetaInsight[93]). Frequentist statistics are commonly used in medical statistics, although criticisms persist. For example, the true definition of a 95% CI is often misunderstood by researchers; it means that if we took 100 different samples then 95% of these CIs (which would all be different) would contain the population value. Conversely, Bayesian statistics use 95% CrIs, which may be more intuitive to understand as they give probabilities that the population value is included within the interval. Although the original Bayes theorem uses prior probabilities, these are often avoided in Bayesian NMA in favour of 'minimally informative priors', although these are not specified in around 50% of published NMAs.[94]

Although Bayesian methods can offer more flexibility, including the range of statistical models that can be specified and estimated via Markov Chain Monte Carlo (MCMC) and related simulation methods, they require greater statistical expertise to conduct.[95] There are specific additional statistics and plots that should be checked to ensure accuracy of the findings.[11] Model fit should be assessed by comparing the posterior mean of the residual deviance to the number of data points. Convergence describes the point when the MCMC method reaches a stable distribution and can be assessed using trace plots and the Gelman-Rubin potential scale reduction factor. When comparing effect estimates derived from frequentist and Bayesian NMA, a simulation study found that they were generally similar in performance.[96] The preference for Bayesian methods is reflected in there being slightly more NMAs in anaesthesia using this method (55% *vs* 40%), although detailed reporting of Bayesian model statistics occurred in only a third of reviews.[9]

Assessing clinical significance in the presence of effect modification

Although the issue of clinical significance is not exclusive to NMA,[6] interpretation of clinical significance in the presence of effect modification is critical. There is frequently a lack of clinical significance reported in reviews of acute pain specific to our previous examples.[97,98] This has the potential to lead to spurious claims and might reduce the clinical use of agents that could have clinical benefit in certain circumstances. This is because whatever the covariate may be, if effect modification exists, then clinical significance varies depending on the value of the covariate. For example, in Fig 2a it can be observed that effects vary based on the covariate. Therefore, quoting a single value does not capture this nuance.[99] Further, the use of Bayesian methods can allow probability statements on whether a treatment surpasses a minimum clinically significant threshold.

Considerations for authors of network meta-analyses

There has been a large increase in NMA publications over recent years. It is therefore important for NMA authors to consider carefully the clinical importance of the topic and results in the review, to incorporate extensive literature searches conducted by qualified and experienced contributors, and to adhere to the highest methodological standards. In addition, the results and conclusions should be presented logically and in a format that is accessible to those reading and reviewing the article. These factors facilitate the translation of NMA results into clinical practice. In the future, artificial intelligence could help automate many of the tasks in an NMA, which is an ongoing area of research.[100]

Conclusions

NMA offers many advantages over pairwise meta-analysis, although it is inherently more complex and has the important additional assumption of transitivity. We present evidence from published studies of violations of this assumption, coupled with inadequate conduct of NMAs in the field of anaesthesia and in the general literature. We present a methodology to help improve these weaknesses and discuss issues for the general reader to help appraise such articles in the literature.

Authors' contributions

Conception, data analysis: BD
Data interpretation, writing manuscript, approving final version, and agreement to be accountable for data: all authors

Declaration of interest

JH is associate editor-in-chief of the *British Journal of Anaesthesia*. The content of this article represents the experience of the authors and a presentation of the reviewed evidence. It does not represent the policy of the *British Journal of Anaesthesia*.

Funding

References

1. Shi J, Gao Y, Ming L, et al. A bibliometric analysis of global research output on network meta-analysis. *BMC Med Inform Decis Mak* 2021; **21**: 144

2. Leucht S, Chaimani A, Cipriani AS, Davis JM, Furukawa TA, Salanti G. Network meta-analyses should be the highest level of evidence in treatment guidelines. *Eur Arch Psychiatry Clin Neurosci* 2016; **266**: 477−80

3. Kanters S, Ford N, Druyts E, Thorlund K, Mills EJ, Bansback N. Use of network meta-analysis in clinical guidelines. *Bull World Health Organ* 2016; **94**: 782

4. Antoniou SA, Mavridis D, Tsokani S, et al. Network meta-analysis as a tool in clinical practice guidelines. *Surg Endosc* 2023; **37**(1): 1−4

5. Cooper N, Germeni E, Freeman SC, et al. New horizons in evidence synthesis for older adults. *Age Ageing* 2023; **52**, afad211

6. Doleman B, Mathiesen O, Jakobsen JC, et al. Methodologies for systematic reviews with meta-analysis of

randomised clinical trials in pain, anaesthesia, and perioperative medicine. *Br J Anaesth* 2021; **126**: 903—11

7. Sivakumar H, Peyton PJ. Poor agreement in significant findings between meta-analyses and subsequent large randomized trials in perioperative medicine. *Br J Anaesth* 2016; **117**: 431—41

8. Conway A, Conway Z, Soalheira K, Sutherland J. High quality of evidence is uncommon in Cochrane systematic reviews in anaesthesia, critical care and emergency medicine. *Eur J Anaesthesiol* 2017; **34**(12): 808—13

9. Sehmbi H, Retter S, Shah UJ, Nguyen D, Martin J, Uppal V. Epidemiological, methodological, and statistical characteristics of network meta-analysis in anaesthesia: a systematic review. *Br J Anaesth* 2023; **130**(3): 272—86

10. Myles PS, Boney O, Botti M, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. *Br J Anaesth* 2018; **120**(4): 705—11

11. Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 2019: F8—12

12. Doleman B, Mathiesen O, Sutton AJ, Cooper NJ, Lund JN, Williams JP. Non-opioid analgesics for the prevention of chronic postsurgical pain: a systematic review and network meta-analysis. *Br J Anaesth* 2023; **130**(6): 719—28

13. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; **23**(20): 3105—24

14. Cooper NJ, Peters J, Lai MC, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health* 2011; **14**(2): 371—80

15. Caldwell DM, Dias S, Welton NJ. Extending treatment networks in health technology assessment: how far should we go? *Value Health* 2015; **18**(5): 673—81

16. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018; **391**(10128): 1357—66

17. Wang L, Huang J, Hu H, Chang X, Xia F. Commonly used antiemetics for prophylaxis of postoperative nausea and vomiting after Caesarean delivery with neuraxial morphine: a network meta-analysis. *Br J Anaesth* 2024; **132**(6): 1274—84

18. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013; **159**(2): 130—7

19. Chaimani A, Caldwell DM, Li T, et al., editors. *Cochrane handbook for systematic reviews of interventions*. Cochrane; 2023 (accessed August 2023). Available from:, version 6.4. www.training.cochrane.org/handbook

20. Papakonstantinou T, Nikolakopoulou A, Egger M, Salanti G. In network meta-analysis, most of the information comes from indirect evidence: empirical study. *J Clin Epidemiol* 2020; **124**: 42—9

21. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: a simulation study. *PloS One* 2011; **6**(1), e16237

22. Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: continuing controversies and recent developments. *Res Synth Methods* 2024; **15**(5): 702—27

23. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network meta-analysis for decision-making*. John Wiley & Sons; 2018

24. Mills EJ, Kanters S, Thorlund K, Chaimani A, Veroniki AA, Ioannidis JP. The effects of excluding treatments from network meta-analyses: survey. *Br Med J* 2013; **347**, f5195

25. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. *NICE DSU Technical Support Document 7: evidence synthesis of treatment efficacy in decision making: a reviewer's checklist* 2012. Available from: http://www.nicedsu.org.uk

26. Efthimiou O, Mavridis D, Debray TP, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med* 2017; **36**(8): 1210—26

27. Toews I, Anglemyer A, Nyirenda JL, et al. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials: a meta-epidemiological study. *Cochrane Database Syst Rev* 2024; **1**. Art. No.: MR000034

28. Li L, Tian J, Tian H, et al. Network meta-analyses could be improved by searching more sources and by involving a librarian. *J Clin Epidemiol* 2014; **67**(9): 1001—7

29. Dobrescu AI, Nussbaumer-Streit B, Klerings I, et al. Restricting evidence syntheses of interventions to English-language publications is a viable methodological shortcut for most medical topics: a systematic review. *J Clin Epidemiol* 2021; **137**: 209—17

30. Hutton B, Catala-Lopez F, Moher D. The PRISMA statement extension for systematic reviews incorporating network meta-analysis: PRISMA-NMA. *Med Clín* 2016; **147**(6): 262—6

31. Veroniki AA, Tsokani S, Zevgiti S, et al. Do reporting guidelines have an impact? Empirical assessment of changes in reporting before and after the PRISMA extension statement for network meta-analysis. *Syst Rev* 2021; **10**: 1—2

32. Veroniki AA, Wong EK, Lunny C, et al. Does type of funding affect reporting in network meta-analysis? A scoping review of network meta-analyses. *Syst Rev* 2023; **12**(1): 81

33. Sideri S, Papageorgiou SN, Eliades T. Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. *J Clin Epidemiol* 2018; **100**: 103—10

34. Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. *Br Med J* 2023; **381**, e074495

35. Nikolakopoulou A, Higgins JP, Papakonstantinou T, et al. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020; **17**(4), e1003082

36. Noori A, Sadeghirad B, Thabane L, Bhandari M, Guyatt GH, Busse JW. The GRADE Working Group and CINeMA approaches provided inconsistent certainty of evidence ratings for a network meta-analysis of opioids for chronic noncancer pain. *J Clin Epidemiol* 2024; **169**, 111276

37. Caldwell DM, Ades AE, Dias S, et al. A threshold analysis assessed the credibility of conclusions from network meta-analysis. *J Clin Epidemiol* 2016; **80**: 68—76

38. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. *PLoS One* 2014; **9**(3), e92508

39. Petropoulou M, Nikolakopoulou A, Veroniki AA, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol* 2017; **82**: 20–8

40. Zarin W, Veroniki AA, Nincic V, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med* 2017; **15**(1): 3

41. Sehmbi H, Retter S, Shah UJ, Nguyen D, Martin J, Uppal V. Methodological and reporting quality assessment of network meta-analyses in anesthesiology: a systematic review and meta-epidemiological study. *Can J Anesth* 2023; **70**(9): 1461–73

42. Doleman B, Sutton AJ, Sherwin M, Lund JN, Williams JP. Baseline morphine consumption may explain between-study heterogeneity in meta-analyses of adjuvant analgesics and improve precision and accuracy of effect estimates. *Anesth Analg* 2018; **126**(2): 648–60

43. Nielsen NI, Kehlet H, Gromov K, et al. High-dose dexamethasone in low pain responders undergoing total knee arthroplasty: a randomised double-blind trial. *Br J Anaesth* 2023; **130**(3): 322–30

44. Nielsen NI, Kehlet H, Gromov K, et al. High-dose steroids in high pain responders undergoing total knee arthroplasty: a randomised double-blind trial. *Br J Anaesth* 2022; **128**(1): 150–8

45. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med* 2013; **11**: 1–8

46. Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; **62**(8): 857–64

47. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 2009; **28**(14): 1861–81

48. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods* 2013; **4**(4): 291–323

49. Papakonstantinou T, Nikolakopoulou A, Higgins JP, Egger M, Salanti G. Cinema: software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Syst Rev* 2020; **16**(1): e1080

50. Béliveau A, Boyne DJ, Slater J, Brenner D, Arora P. BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network Meta-analyses. *BMC Med Res Methodol* 2019; **19**: 1–3

51. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; **29**(7-8): 932–44

52. Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol* 2012; **12**: 1–4

53. Veroniki AA, Mavridis D, Higgins JP, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. *BMC Med Res Methodol* 2014; **14**: 1–2

54. Freeman SC, Fisher D, White IR, Auperin A, Carpenter JR. Identifying inconsistency in network meta-analysis: is the net heat plot a reliable method? *Stat Med* 2019; **38**(29): 5547–64

55. Jackson D, Barrett JK, Rice S, White IR, Higgins JP. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat Med* 2014; **33**(21): 3639–54

56. Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. *Int J Epidemiol* 2013; **42**: 332–45

57. Veroniki AA, Tsokani S, White IR, et al. Prevalence of evidence of inconsistency and its association with network structural characteristics in 201 published networks of interventions. *BMC Med Res Methodol* 2021; **21**: 1–10

58. Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *Br Med J* 2011; **343**, d4909

59. Spineli LM, Kalyvas C, Yepes-Nuñez JJ, et al. Low awareness of the transitivity assumption in complex networks of interventions: a systematic survey from 721 network meta-analyses. *BMC Med* 2024; **22**(1): 1–7

60. Gao Y, Ge L, Ma X, Shen X, Liu M, Tian J. Improvement needed in the network geometry and inconsistency of Cochrane network meta-analyses: a cross-sectional survey. *J Clin Epidemiol* 2019; **113**: 214–27

61. Kovic B, Zoratti MJ, Michalopoulos S, Silvestre C, Thorlund K, Thabane L. Deficiencies in addressing effect modification in network meta-analyses: a meta-epidemiological survey. *J Clin Epidemiol* 2017; **88**: 47–56

62. Zufferey PJ, Chaux R, Lachaud PA, Capdevila X, Lanoiselée J, Ollier E. Dose–response relationships of intravenous and perineural dexamethasone as adjuvants to peripheral nerve blocks: a systematic review and model-based network meta-analysis. *Br J Anaesth* 2024; **132**(5): 1122–32

63. Freeman SC, Scott NW, Powell R, Johnston M, Sutton AJ, Cooper NJ. Component network meta-analysis identifies the most effective components of psychological preparation for adults undergoing surgery under general anesthesia. *J Clin Epidemiol* 2018; **98**: 105–16

64. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; **64**(2): 163–71

65. Hedin RJ, Umberham BA, Detweiler BN, Kollmorgen L, Vassar M. Publication bias and nonreporting found in majority of systematic reviews and meta-analyses in anesthesiology journals. *Anesth Analg* 2016; **123**(4): 1018–25

66. Chong SW, Collins NF, Wu CY, Liskaser GM, Peyton PJ. The relationship between study findings and publication outcome in anesthesia research: a retrospective observational study examining publication bias. *Can J Anesth* 2016; **63**(6): 682

67. Chong SW, Imberger G, Karahalios A, et al. Trial registration of abstracts from the American Society of Anesthesiologists Meetings 2010–2016: a review of prospective trial registration and selective outcome reporting. *PLoS One* 2022; **17**(7), e0270841

68. Chong SW, Imberger G, Karahalios A, et al. The relationship between study findings and publication outcome in anesthesia research following

implementation of mandatory trial registration: a systematic review of publication bias. *PLoS One* 2023; **18**(5), e0282839

69. Jones PM, Chow JT, Arango MF, et al. Comparison of registered and reported outcomes in randomized clinical trials published in anesthesiology journals. *Anesth Analg* 2017; **125**(4): 1292–300

70. Biocic M, Fidahic M, Cikes K, Puljak L. Comparison of information sources used in Cochrane and non-Cochrane systematic reviews: a case study in the field of anesthesiology and pain. *Res Synth Methods* 2019; **10**(4): 597–605

71. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. *Stat Med* 2014; **33**(30): 5399–412

72. Chiocchia V, Nikolakopoulou A, Higgins JP, et al. ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. *BMC Med* 2021; **19**: 1–3

73. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005; **58**: 894e901

74. Thorlund K, Mills EJ. Sample size and power considerations in network meta-analysis. *Syst Rev* 2012; **1**: 1–3

75. Del Re AC, Spielmans GI, Flückiger C, Wampold BE. Efficacy of new generation antidepressants: differences seem illusory. *PLoS One* 2013; **8**(6), e63509

76. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J* 1986; **292**(6522): 746–50

77. Wang E, Belley-Côté EP, Young J, et al. Effect of perioperative benzodiazepine use on intraoperative awareness and postoperative delirium: a systematic review and meta-analysis of randomised controlled trials and observational studies. *Br J Anaesth* 2023; **131**(2): 302–13

78. Li H, Shih MC, Song CJ, Tu YK. Bias propagation in network meta-analysis models. *Res Synth Methods* 2023; **14**(2): 247–65

79. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Ann Int Med* 2008; **148**(7): 544–53

80. Batson S, Score R, Sutton AJ. Three-dimensional evidence network plot system: covariate imbalances and effects in network meta-analysis explored using a new software tool. *J Clin Epidemiol* 2017; **86**: 182–95

81. Donegan S, Dias S, Tudur-Smith C, Marinho V, Welton NJ. Graphs of study contributions and covariate distributions for network meta-regression. *Res Synth Methods* 2018; **9**(2): 243–60

82. Papakonstantinou T, Nikolakopoulou A, Rücker G, et al. Estimating the contribution of studies in network meta-analysis: paths, flows and streams. *F1000Research* 2018; **7**: 610

83. Seo M, Furukawa TA, Veroniki AA, et al. The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes. *Res Synth Methods* 2021; **12**(1): 86–95

84. Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel presentational approaches were developed for reporting network meta-analysis. *J Clin Epidemiol* 2014; **67**(6): 672–80

85. Da Costa BR, Pereira TV, Saadat P, et al. Effectiveness and safety of non-steroidal anti-inflammatory drugs and opioid treatment for knee and hip osteoarthritis: network meta-analysis. *Br Med J* 2021; **375**, n2321

86. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015; **15**: 1–9

87. Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in treatment rankings: reanalysis of network meta-analyses of randomized trials. *Ann Int Med* 2016; **164**(10): 666–73

88. Veroniki AA, Straus SE, Rücker G, Tricco AC. Is providing uncertainty intervals in treatment ranking helpful in a network meta-analysis? *J Clin Epidemiol* 2018; **100**: 122–9

89. Nevill CR, Cooper NJ, Sutton AJ. A multifaceted graphical display, including treatment ranking, was developed to aid interpretation of network meta-analysis. *J Clin Epidemiol* 2023; **157**: 83–91

90. Shim SR, Kim SJ, Lee J, Rücker G. Network meta-analysis: application and practice using R software. *Epidemiol Health* 2019; **41**, e2019013

91. Slater J, Beliveau A, Boyne DJ, Brenner DR, Arora P. PNS33 BUGSNET: a new comprehensive R package for network meta-analysis. *Value Health* 2019; **22**: S292

92. Balduzzi S, Rücker G, Nikolakopoulou A, et al. netmeta: an r package for network meta-analysis using frequentist methods. *J Stat Softw* 2023; **106**: 1–40

93. Owen RK, Bradbury N, Xin Y, Cooper N, Sutton A. Meta-Insight: an interactive web-based tool for analyzing, interrogating, and visualizing network meta-analyses using R-shiny and netmeta. *Res Synth Methods* 2019; **10**: 569–81

94. Rosenberger KJ, Xing A, Murad MH, Chu H, Lin L. Prior choices of between-study heterogeneity in contemporary Bayesian network meta-analyses: an empirical study. *J Gen Int Med* 2021; **36**: 1049–57

95. Sadeghirad B, Foroutan F, Zoratti MJ, et al. Theory and practice of Bayesian and frequentist frameworks for network meta-analysis. *BMJ Evid Based Med* 2023; **28**(3): 204–9

96. Seide SE, Jensen K, Kieser M. A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res Synth Methods* 2020; **11**(3): 363–78

97. Hussain N, Brull R, Weber L, et al. The analgesic effectiveness of perioperative lidocaine infusions for acute and chronic persistent postsurgical pain in patients undergoing breast cancer surgery: a systematic review and meta-analysis. *Br J Anaesth* 2024; **132**(3): 575–87

98. Verret M, Lauzier F, Zarychanski R, et al. Perioperative use of gabapentinoids for the management of postoperative acute pain: a systematic review and meta-analysis. *Anesthesiology* 2020; **133**(2): 265–79

99. Doleman B, Lund JN, Williams JP. Clinically significant reductions in morphine consumption need to take account of baseline risk: presentation of a novel meta-analysis methodology. *Br J Anaesth* 2018; **120**(2): 414–5

100. Reason T, Benbow E, Langham J, Gimblett A, Klijn SL, Malcolm B. Artificial intelligence to automate network meta-analyses: four case studies to evaluate the potential application of large language models. *Pharmacoecon Open* 2024; **8**(2): 205–20