ELSEVIER

Journal of Clinical Epidemiology 98 (2018) 62−69

Journal of Clinical Epidemiology

REVIEW

Using decision thresholds for ranking treatments in network meta-analysis results in more informative rankings

Romina Brignardello-Petersen[a,b,*], Bradley C. Johnston[c,d,e], Alejandro R. Jadad[c,f,g], George Tomlinson[c,h]

[a]Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street W, Hamilton, Ontario L8S 4L8, Canada
[b]Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Sergio Livingstone 943, Independencia, Santiago, Chile
[c]Institute of Health Policy, Management and Evaluation, University of Toronto, 155 College Street, Toronto, Ontario M5T 3M6, Canada
[d]Child Health Evaluative Sciences, Hospital for Sick Children Research Institute, Peter Gilgan Centre for Research and Learning, 686 Bay Street, Rm 11.9859 West, Toronto, Ontario M5G 0A4, Canada
[e]Department of Anesthesia and Pain Medicine, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada
[f]Institute for Global Health Equity and Innovation, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, Ontario M5T 3M6, Canada
[g]Centre for global eHealth Innovation, R Fraser Elliot Building, 190 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada
[h]Department of Medicine, University Health Network and Mt Sinai Hospital, Toronto, Eaton North, 10th floor, room 235, 200 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada

Accepted 1 February 2018; Published online 14 February 2018

Abstract

Objectives: To evaluate how the rank probabilities obtained from network meta-analysis (NMA) change with the use of increasingly stringent criteria for the relative effect comparing two treatments which ranks one treatment better than the other.

Study Design and Setting: Systematic survey and reanalysis of published data. We included all systematic reviews (SRs) with NMA from the field of cardiovascular medicine that had trial-level data available, published in Medline up to February 2015. We reran all the NMAs and determined the probabilities of each treatment being the best. For the best treatment, we examined the effect on these probabilities of varying, what we call the decision threshold, the relative effect required to declare two treatments different.

Results: We included 14 SRs, having a median of 20 randomized trials and 9 treatments. The best treatments had probabilities of being best that ranged from 38% to 85.3%. The effect of changing the decision thresholds on the probability of a treatment being best varied substantially across reviews, with relatively little decrease (~20 percentage points) in some settings but a decline to near 0% in others.

Conclusion: Rank probabilities can be fragile to increases in the decision threshold used to claim that one treatment is more effective than another. Including these thresholds into the calculation of rankings may aid their interpretation and use in clinical practice.

Keywords: Decision threshold; Network meta-analysis; Rank probabilities; Results interpretation; Rankings; Systematic survey

1. Introduction

In many areas of medicine, multiple treatment options are available for the treatment of a particular condition. Over time, clinical trials appear that each compare two or perhaps three of these treatments. The resulting body of evidence can be seen as a network where any two treatments may have direct comparisons (from head-to-head trials), indirect comparisons (through a common comparator or sequence of comparators), or a mixture of both direct and indirect comparisons. Network meta-analysis (NMA), also called mixed treatment comparison meta-analysis, is a statistical technique used to combine the results of such a series of trials and estimate the relative effectiveness of all the interventions included in the network [1]. NMA has become very popular in recent years, with fewer than 10 of these studies published each year from 2003 to 2008, but approximately 85 published in 2013 [2]. The number of hits when searching for NMA and NMA-related terms in databases such as PubMed has increased steadily [2], a

Conflict of interest: None of the authors of this study has any conflict of interest to declare.

The code we used is available as an online appendix (Appendix 6).

* Corresponding author. Tel.: +1 905-531-7013; fax: +1 (905)-524-3841.

E-mail address: brignarr@mcmaster.ca (R. Brignardello-Petersen).

https://doi.org/10.1016/j.jclinepi.2018.02.008

What is new?

Key findings

- The probability that a treatment in a network meta-analysis (NMA) is the best one may be changed substantially when a decision threshold is incorporated in the calculation of this probability.

What this adds to what was known?

- Because decision thresholds can have an important effect on the rank probabilities and rankings from NMA, their use could change the conclusions drawn by reviewers and clinicians when they use the evidence in NMA.

What is the implication and what should change now?

- To reduce the chance of reaching misleading conclusions based on rankings, systematic reviewers should consider decision thresholds explicitly and transparently when conducting NMA.

sign that this technique is being developed, studied, and discussed more often.

The large majority of NMAs use Bayesian methods for fitting statistical models to the data in the network. An appealing feature of the Bayesian approach to NMA, frequently highlighted by its proponents, is the ability to obtain the probabilities that each treatment is the best (or second best, or worse) for a specific outcome; these are known as the rank probabilities and allow the treatments to be ranked from best to worst [3–6]. The ability to generate rank probabilities is a consequence of the use of Markov Chain Monte Carlo (MCMC) methods for model fitting. In each MCMC iteration, every treatment can be ranked based on its relative effect with respect to some chosen reference treatment. Then, the probability of each treatment being the best (or second best, and so forth) is calculated as the proportion of MCMC iterations in which each treatment was ranked first. The treatment with the highest probability of being best is ranked as the best treatment; whereas the ranking with the second highest rank probability of being the best is ranked as the second best treatment, and so on.

Rankings obtained from NMA are potentially misleading [7,8] as they emphasize the probability of a treatment being the best, without any consideration of the clinical importance of the size of the treatment effects. Because the standard approach to obtain the rankings considers any nonzero treatment effect to be relevant, the standard rankings do not formally take into account how much better a treatment is compared with the next best treatments [8,9]. This problem is highlighted by a recent article [10] that cites an odds ratio (OR) from an NMA of 1.42 (95% Credible Interval (CrI): 0.89-2.25) comparing telaprevir to boceprevir for sustained virological response in hepatitis C. Since the CrI includes 1, this is described as "no statistical difference", but the probability of telaprevir being the best treatment in the NMA is 93%. The authors go on to say "this 93% probability provides a misleadingly strong endorsement for telaprevir. One may wish to know the probability that telaprevir is better than boceprevir by a clinically important margin." The advice is to interpret the probabilities "with caution." To date, research in this field offers no further insight into how the choice of a clinically important margin may affect the rank probabilities, and there is no standard approach for calculation of rankings in light of a stated clinically important margin for a treatment effect. We term this clinically important margin (which some call "minimally important difference") a decision threshold.

Our aim was to explore to what extent the rank probability of the best treatment in an NMA changed when using decision thresholds (e.g., an OR less than 1, an OR than 0.9, etc.) for declaring that one treatment is better than another. We did this through a systematic survey and reanalysis of all the NMAs that had trial-level data available in the field of cardiovascular medicine, the medical specialty with the most published NMAs [2].

2. Methods

2.1. Eligibility criteria

We included a study if it met the following eligibility criteria: (1) it was a systematic review (SR) of randomized controlled trials (RCTs), defined as a review that performed a systematic search of RCTs in at least one electronic database, and where authors selected articles based on specific and explicit selection criteria; (2) the question of interest concerned more than two interventions (any drugs, administered by any route, or stents), which should be explicit in the aim, methods, or results of the review; (3) the study used NMA as the primary or secondary method of analysis, defined as the use of a single statistical model on an evidence network that involves two or more RCTs and at least three interventions; (4) it addressed a question in the field of cardiovascular medicine; (5) it assessed a dichotomous primary outcome (as specified by the authors) with a pooled estimate reported as an relative risk, or OR; and (6) it had trial-level data available in either the manuscript or online supplements for all the interventions included in the network. We excluded SRs with NMAs where patients had a cardiovascular disease, yet both the interventions and outcomes were relevant to another medical field (e.g., diabetes) and studies that were not published in the English language.

2.2. Study searching and selection

We constructed an electronic search for OVID Medline, which was run from the database's inception to February

2015 (see Appendix 1). After removing duplicates, we screened the titles and abstracts of all the citations and gathered the full-text screening of all the references that seemed relevant. All these articles were screened in full-text to confirm eligibility. Two reviewers screened articles at both stages independently, and in duplicate. All final decisions were reached through consensus, and a third reviewer provide advice in those cases were there was doubt regarding some of the criteria (in particular, whether the method of analysis performed was indeed an NMA).

2.3. Data abstraction

We abstracted data for the primary studies included in each NMA. For this, we used the data reported by the SR authors in the tables of the articles or, when necessary, online supplementary material. When there was inconsistency between the data presented in figures and tables, or when information was missing (such as the number of events for a specific trial), we referred to the primary study to find the information. More details with regards to the data abstraction process can be found in Appendix 2.

2.4. Data analysis and outcomes

We reran the NMA for each of the SRs, using the abstracted data. We used the OR as the measure of effect and calculated the usual rank probabilities from the proportion of MCMC iterations in which each of the treatments was ranked first (or second, or third, and so forth) when comparing all the treatments to a reference treatment by means of the OR. These rankings were assigned based on any nonzero difference between treatments, that is, an OR for the relative effect between treatments that was less than 1 for harm (bad outcomes) and greater than 1 for benefit (good outcomes).

Subsequently, we calculated the rank probabilities using an increasingly stringent set of thresholds for superiority (ORs starting at 1 and decreasing to 0.6 for bad outcomes, and ORs starting at 1 and increasing to 1.67 for good outcomes). Table 1 shows how the adjusted rankings for a treatment are calculated, using a hypothetical single MCMC sample of ORs from an NMA comparing mortality between three treatments, A, B, and C, where the OR comparing treatments A and B is 0.75 and the OR comparing treatments A and C is 0.9. Treatment A is classified as better than both B and C at the usual threshold of OR = 1; it is the single best treatment. But at a threshold of 0.9, meaning that treatments having a relative OR of 0.9 or larger are not different to a clinically importantly degree, we do not rank A better than C; at this threshold, treatment A is merely one of the top two treatments. Once the threshold reaches an OR of 0.7, A is no better than B or C; therefore, we conclude only that it is one of the top three treatments. This same procedure can be applied to each of the treatments A, B, and C. For each of the thresholds listed previously, we computed the revised rank probabilities as the proportion of MCMC samples where a treatment was ranked 1 and the proportion where it was ranked either 1 (best) or "1 or 2" (one of the top two treatments) and so on.

We used the *gemtc* package [11] in R [12] to carry out random-effects Bayesian hierarchical consistency models with uninformative priors [1]. We used MCMC with four parallel chains for assessment of convergence with an adaptation phase of 5,000 samples and 20,000 burn-in iterations, at which point we checked convergence of the models using the Gelman-Rubin statistic [13]. If models had not converged, we ran further iterations until convergence was reached. If convergence was not reached after 400,000 iterations, we explored the cause. After identifying the parameters and RCTs that caused the issue, we either excluded the RCT if it had no events in any arm (because it was not providing any extra information for the estimation of the parameter) or added one event to both arms if one arm had 0 events and the other had 1 or more. We explored if any convergence issues remained after doing this and if so, reported this explicitly in the results.

The probabilities of each treatment being the best under varying thresholds were obtained using our modification of the *rank.probability* built-in function in the *gemtc* package, and the revised rankings were obtained from these probabilities. We report the sensitivity to the threshold for only the treatment ranked first using the default threshold OR of 1.

3. Results

The search resulted in 975 references, from which 131 articles were screened in full-text and 14 were included. The reasons for exclusions of articles were not performing

Table 1. Adjusted rankings calculation

Values (OR) on MCMC iteration	**$OR_{(A\ vs.\ B)} = 0.75$**	**$OR_{(A\ vs.\ C)} = 0.91$**		
Threshold for declaring superiority	**A better than B?**	**A better than C?**	**Number of treatments A is superior to**	**Rank of A**
OR < 1.0	Yes	Yes	2	1
OR < 0.9	Yes	No	1	1 or 2
OR < 0.8	Yes	No	1	1 or 2
OR < 0.7	No	No	0	1, 2, or 3
OR < 0.6	No	No	0	1, 2, or 3

Abbreviations: MCMC, Markov Chain Monte Carlo; OR, odds ratio.

an SR (18/117), asking a question about only two interventions (26/117), using a method of meta-analysis that was not an NMA (24/117), having an outcome that was not dichotomous or reported as relative risk or OR (7/117), and not providing data for rerunning the NMA (42/117). The SRs included were published between 2003 and 2015 and covered a wide range of patients, interventions, and outcomes from the field of cardiovascular medicine (see Table 2). These NMAs for the primary outcomes included a median of 20 RCTs (range 11-57) and nine treatments (range 4-12). The geometry of all networks is shown in Appendix 3.

3.1. Rank probabilities of the best two treatments when using a threshold OR of 1

The NMAs rarely had convincing evidence about the supremacy of any particular intervention: the best-ranked treatments had a mean probability of 58.5% (range 34.8-85.3%) of being the best treatment. For treatments ranking second best, the mean probability of being the best treatment was 20.5% (range 10.3-32.7%; Table 3). The best and second best treatments had a mean total probability of being in the top two of 79.0% (ranging from 58.2% in an NMA with 11 treatments [24] to 99.2% in a network with five treatments [20]).

3.2. Sensitivity to the threshold used declare two treatments different

NMAs behaved differently when we increased the stringency of the threshold OR for declaring a difference and recalculating rank probability for the best treatment. Although in some of them the change in rank probability was small [14,17,21]; in others, the probability decreased rapidly, reaching low values [16,22,24,26]. Fig. 1 shows the probabilities of the best treatment being the best or one of the top two treatments vs. an increasingly stringent threshold OR.

The overall change in the probability of the best treatment being the best, and the slope of this change against the threshold depended on the specific NMA. There were NMAs in which the best treatment had small changes when increasing the OR threshold, and the range in the total

Table 2. Main characteristics of the included SRs

				Numbers		
Study	Patients	Interventions	Primary outcome	RCTs	Treatments	Direct comparisons
Bash, 2012 [14]	Atrial fibrillation	Interventions to achieve cardioversion	Successful cardioversion	20	11	13
Castellucci, 2013 [15]	Venous thromboembolism	Antiplatelet or oral anticoagulant	Recurrent venous thromboembolism	11	9	11
Coleman, 2008 [16]	Undergoing treatment with antihypertensives	Antihypertensives	Cancer	27	6	10
Cooper, 2006 [17]	Nonrheumatic atrial fibrillation	Stroke prevention agents	Stroke	19	9	14
Dogliotti, 2014 [18]	Atrial fibrillation	Antithrombotics	Stroke	20	8	11
Dooley, 2014 [19]	Hospitalized patients	Low-molecular-weight heparins	Mortality	14	9	11
Harenberg, 2012 [20]	Undergoing total hip or knee replacement	New oral anticoagulants	Venous thromboembolism	16	5	5
Landoni, 2013 [21]	Undergoing cardiac surgery	Anesthetic drugs	Mortality	36	4	5
Navarese, 2013 [22]	Undergoing treatment with statins	Statins	Diabetes	17	12	14
Phung, 2011 [23]	Hospitalized, at risk of venous thromboembolism	Thromboprophylaxis	Deep venous thrombosis	13	4	5
Psaty, 2003 [24]	Undergoing treatment with anihypertensives	Antihypertensives	Coronary heart disease	27	11	18
Roskell, 2010 [25]	Atrial fibrillation	Anticoagulants	Stroke	13	12	17
Sciarretta, 2011 [26]	Hypertension	Antihypertensives	Heart failure	26	8	16
Wu, 2013 [27]	Diabetes	Antihypertensives	Mortality	57	8	14

Abbreviations: RCTs, randomized controlled trials.

Table 3. Treatments ranked first and second in each of the NMAs and their rank probabilities

Study	Best treatment	Probability (%)	Second best treatment	Probability (%)
Bash, 2012 [14]	Vernakalant iv	85.3	Flecainide oral	10.3
Castellucci, 2013 [15]	Standard dose vitamin K antagonist	57.1	Dabigatran 150 mg tid	18.5
Coleman, 2008 [16]	Diuretics	38.3	Beta-blockers	23.2
Cooper, 2006 [17]	Alternate day aspirin	63.0	Low-dose warfarin	11.0
Dooley, 2014 [19]	Fondaparinux	64.7	Enoxaparin 40 mg	15.3
Dogliotti, 2014 [18]	Dabigatran 150 mg	64.6	Rivaroxaban	19.8
Harenberg, 2012 [20]	Rivaroxaban	66.5	Apixaban	32.7
Landoni, 2013 [21]	Desflurane	67.4	Isoflurane	31.5
Navarese 2013 [22]	Pravastatin 20 mg	40.6	Lovastatin	27.8
Phung, 2011 [23]	Unfractionated heparin bid	75.6	Low-molecular-weight heparin	15.0
Psaty, 2003 [24]	Beta-blockers/diuretics	38.0	Ace inhibitors	20.2
Roskell, 2010 [25]	Ximelagatran	34.8	Dabigatran 150 mg tid	32.3
Sciarretta, 2011 [26]	ACE inhibitors	49.3	Angiotensin receptor blockers	15.9
Wu, 2013[a] [27]	ACE inhibitor + calcium channel blocker	73.4	Diuretics	13.6

Abbreviations: ACE, angiotensin-converting-enzyme; NMA, network meta-analysis.

[a] There were convergence issues when performing this NMA; however, these were observed in the estimation of treatment effects different from the best treatment.

probability change was less than 20 percentage points [14,17,21]. In these cases, the best treatments started with probabilities higher than 60%. In some NMAs, the change in rank probability was moderate, less than 35 percentage points [19,23,25,27], whereas in others, it was up to 50 percentage points [15,18,20]. Finally, there were NMAs in which the rank probabilities changed from approximately 40% at the threshold of OR = 1 to very low probabilities of being the best treatments when the threshold was OR 0.8 and almost 0% at the threshold of OR = 0.6 [16,22,24,26].

We observed similar patterns when quantifying the probability of each treatment being the best or second best (Appendix 4). Appendix 5 shows the variation in rank probabilities for each of the NMAs separately.

There was no apparent relationship between the size of the network, in terms of trials or treatments, and the sensitivity of the rankings to these changing thresholds.

4. Discussion

In this study, we performed a systematic survey of NMAs in cardiovascular medicine that reported trial-level data. With these data, we explored changes in rank probabilities that resulted from the use of more stringent thresholds (i.e., OR of 0.9-0.6 for bad outcomes) for declaring that two treatments had a different effect. As expected, more stringent thresholds decreased the size of the probability that a given treatment was best, but the magnitude of this change depended on the specific case.

The rankings are among one of the most cited advantages of NMA [3,5,28]. The attractiveness of the rankings lies in their simplicity to illustrate which treatment is the best for a specific outcome, a notion that is easy to understand, and has great appeal in cases in which alternative treatments are numerous. The summaries of comparative effectiveness based on rankings, increase the potential to facilitate the decision-making process [28,29]. Nevertheless, rankings can be misleading since they do not incorporate information about the magnitude of differences in treatment effects [7–9]. This study explores the relationship between size of the treatment effect (or decision threshold to claim that one treatment is better than the other) and the subsequent rank probabilities. We calculated the rank probabilities using the conventional decision threshold of a relative effect of 1 (i.e., one treatment is better than the other if the OR comparing the two treatments for a bad outcome is <1 or >1 for a good outcome), and using wide range of decision thresholds: relative effects from 0.9 to 0.6 for bad outcomes and from 1.1 to 1.6 for good outcomes. For any particular NMA, the choice of the appropriate threshold should depend on the cost, invasiveness, and risk of the treatment and the seriousness of the outcome. Although the ranges above cover the most likely values of decision thresholds in practice, in any particular NMA, the range could be narrowed based on subject area knowledge. For example, if the outcome is mortality, the threshold may be around OR = 0.9; whereas an expensive intervention with a less severe outcome may require a threshold from OR 0.7 to 0.8. Treatment effects used to design the primary RCTs in the NMA, for example, could be a source of clinically relevant ranges of threshold values.

It was predictable that as we applied more stringent criteria for declaring two treatments to have a different effect, the rank probabilities of the best treatment decreased.

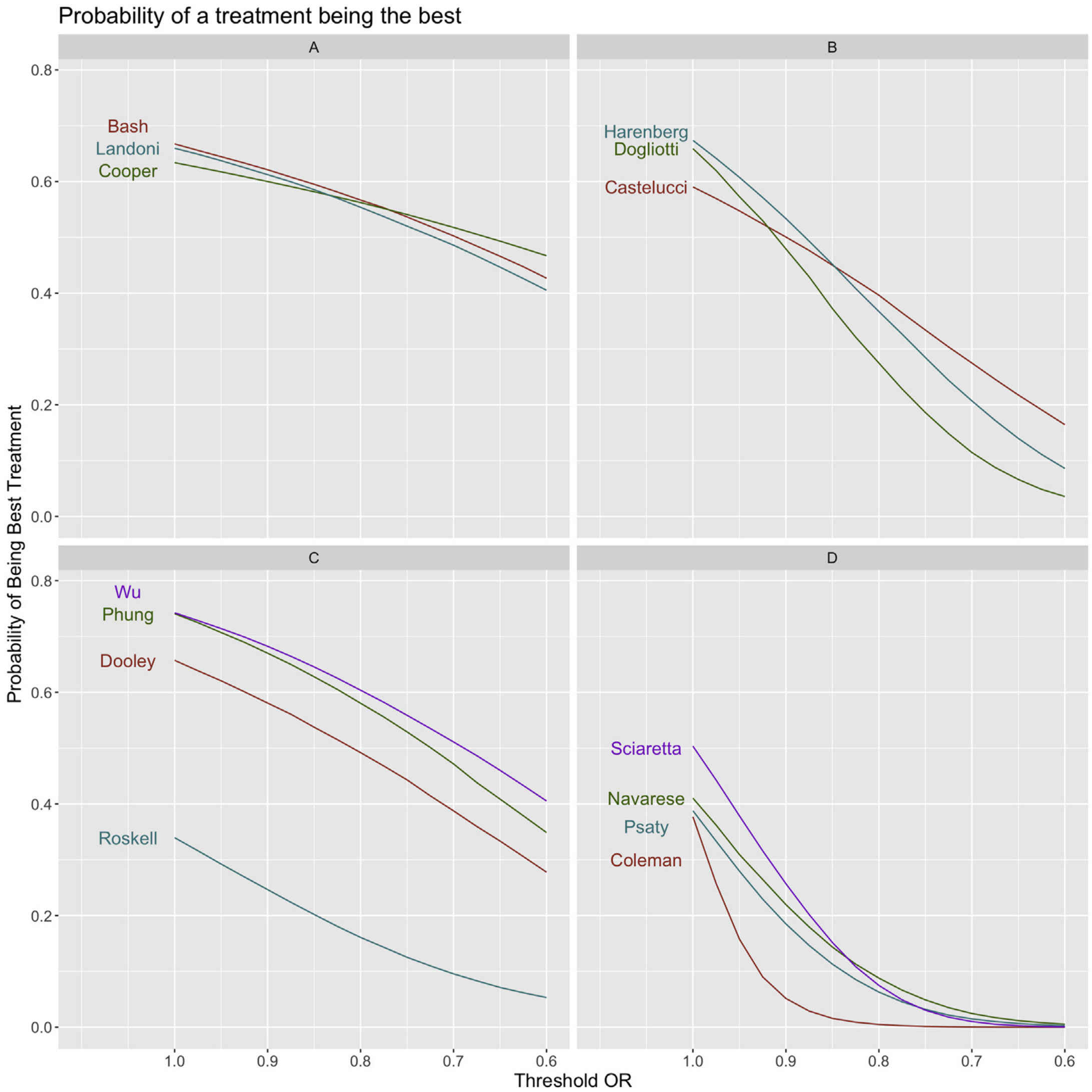

Fig. 1. Change in the probabilities of the best treatment when increasing the OR thresholds to calculate these probabilities. (A) small change; (B) constant moderate decrease in probabilities; (C) constant large decrease in probabilities; and (D) rapid decrease to very small probabilities. In Bash et al., the outcomes were a desirable outcome, and therefore, we present an OR that is the reciprocal of the relevant OR for an increase in the odds of the outcome. OR, odds ratio.

However, the magnitude and pattern of the decrease was specific to each NMA. We observed cases with small decreases in probability across increasing thresholds at one extreme, and cases with rapid decreases in probability that went down to 0% in others. This supports the notion that the interpretation of the rankings must be accompanied by a careful interpretation of the pairwise estimates comparing the best treatment with the other treatments [8,10]. One such careful interpretation is the recalculation of the rankings using a set of thresholds relevant to the particular NMA, an approach that could become a standard part of the analyses of networks of evidence. This method can be easily applied to not just the best treatment but also the second best, third best, and so on.

To our knowledge, and despite the fact that potential issues with the use of rankings have been raised [8,9], there is only one systematic survey that has explored this in more depth [30]. The focus there, however, was the effect of excluding treatments from the network on the identification of the best and the three top treatments. There was no consideration of size of the treatment effect.

In this study, we chose to focus on the effects of the interventions on a particular outcome in the context of an SR whose role is to provide information about the effects of interventions. Therefore, we applied the same decision threshold to all pairwise comparisons within a network. We acknowledge that this is a simplification, and that for decision-making, there is a trade-off between all desirable and undesirable consequences of an intervention, which can result in different decision thresholds for different pairwise comparisons. It is possible to adapt our approach to use different decision thresholds for different pairwise comparisons and get more contextualized answer with regard to which is more likely to be the best treatment.

Our study has several strengths. As a first attempt to explore the issue of size of the treatment effect around the rankings in NMA, we designed our study to follow systematic and feasible methods. We chose NMAs from the field of cardiovascular medicine because this is the field with the most published NMAs [2]. Since the outcomes studied in these NMAs are those common in this field, our results could be applicable to the majority of the NMAs in this area. We chose to use published NMAs with available data as our source of data to explore how the rank probabilities would be affected by applying thresholds for superiority in real scenarios, as opposed to simulated data sets. We did not try to replicate the results from the published studies. We acknowledge, therefore, that our analyses are only an exercise and that our results should not be used to make clinical decisions.

Our study also has limitations. One of our inclusion criteria was the availability of primary data. In addition, we only searched from NMAs published in one electronic database in one clinical area. However, there is no reason to doubt the credibility of the finding describing the changing relationships between decision thresholds treatment effects and revised rank probabilities because they are based only on SR with available primary data that were published in a journal indexed in this database. A second limitation is related to the reporting of the trial-level data in the NMAs. In a few cases, some small amount of data processing was needed so that the reported data could be used in the NMA. These issues may explain any differences between the results reported in the original SR and the ones we obtained when rerunning the NMAs, but they do not affect our overall conclusions regarding the robustness of rankings. Third, owing to computational difficulties that led to nonconvergence after 400,000 iterations, we had to exclude RCTs in which the number of events in both arms was zero. Our results, therefore, may not completely apply to scenarios in which there are multiple RCTs with zero events. A final limitation is that the NMAs here did not tend to find any treatments with overwhelmingly high probabilities of being best; it is possible that where there is a clearly best treatment, rank probabilities are less fragile.

Our findings have several implications for systematic reviewers using NMAs, for clinicians who may inform their practice using SRs that report NMAs, and for further methodological research in NMA. First, authors of SRs that use NMAs, peer reviewers, and journal editors should be careful when interpreting rank probabilities and rankings in their articles. They must acknowledge that although this information may be useful and attractive, they must draw conclusions about treatment effects that also take into account the size of the difference in effectiveness in a transparent manner. Authors of SRs using NMAs focused on dichotomous outcomes could also consider establishing a decision threshold for each outcome that they are assessing, and estimate the rank probabilities and rankings based on those. Secondly, our results highlight the need to interpret rankings and rank probabilities with caution when using SRs with NMA to inform clinical practice. Although rankings are highly attractive, they have the potential to be misleading if they are used as a stand-alone piece of information. If our approach using decision thresholds is not considered in the calculation of the rank probabilities, literature users are encouraged to consider both the relative effects of the pairwise comparisons and the rankings to make their conclusions with regard to the effectiveness of a set of treatments, even if this requires a bigger effort on their part. Finally, methodologists should investigate how changing the decision thresholds affects the rank probabilities and ranking in SRs from other medical fields, and with NMAs assessing continuous outcomes.

We have used the OR as an example of how decision thresholds can be used to obtain modified rankings. It is also relatively straightforward within a Bayesian NMA to obtain predicted response probabilities for each treatment and to use thresholds for absolute risk differences to obtain the rankings. Similarly, with continuous outcomes, mean differences could be compared with accepted minimal important difference estimates for health status measures to obtain modified rankings.

5. Conclusions

Rank probabilities can be fragile to changes in the decision threshold used to claim that one treatment is more effective than another. The probability that a given treatment is best can decrease, sometimes dramatically, with a more stringent threshold. This highlights the need for reporting, interpreting, and using rankings together with the pairwise comparison estimates. Modifying the way in which rank probabilities and rankings are estimated by including a sensitivity analysis using different thresholds would facilitate their interpretation and use. This sensitivity analysis would also avoid the need to combine in some unspecified way the two pieces of information, the rankings and the pairwise estimates.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2018.02.008.

References

[1] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105−24.

[2] Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Cabrini L, Zangrillo A, et al. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. Heart Lung Vessel 2015;7:133−42.

[3] Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods 2012;3:80—97.

[4] Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008;26:753—67.

[5] Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, et al. GetReal in network meta-analysis: a review of the methodology. Res Synth Methods 2016;7:236—63.

[6] Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008;11:956—64.

[7] Neumann I, Brignardello-Petersen R, Guyatt G, ACP Journal Club. Review: novel oral anticoagulants reduce stroke more than ASA in nonvalvular atrial fibrillation. Ann Intern Med 2014;160:Jc3.

[8] Brignardello-Petersen R, Rochwerg B, Guyatt GH. What is a network meta-analysis and how can we use it to inform clinical practice? Pol Arch Med Wewn 2014;124:659—60.

[9] Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. BMJ 2013;346:f2914.

[10] Mills EJ, Ioannidis JP, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012;308:1246—53.

[11] van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Res Synth Methods 2012;3:285—99.

[12] R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

[13] Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. J Comput Graph Stat 1998;7:434—45.

[14] Bash LD, Buono JL, Davies GM, Martin A, Fahrbach K, Phatak H, et al. Systematic review and meta-analysis of the efficacy of cardioversion by vernakalant and comparators in patients with atrial fibrillation. Cardiovasc Drugs Ther 2012;26:167—79.

[15] Castellucci LA, Cameron C, Le Gal G, Rodger MA, Coyle D, Wells PS, et al. Efficacy and safety outcomes of oral anticoagulants and antiplatelet drugs in the secondary prevention of venous thromboembolism: systematic review and network meta-analysis. BMJ 2013;347:f5133.

[16] Coleman CI, Baker WL, Kluger J, White CM. Antihypertensive medication and their impact on cancer incidence: a mixed treatment comparison meta-analysis of randomized controlled trials. J Hypertens 2008;26:622—9.

[17] Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. Arch Intern Med 2006;166:1269—75.

[18] Dogliotti A, Paolasso E, Giugliano RP. Current and new oral antithrombotics in non-valvular atrial fibrillation: a network meta-analysis of 79 808 patients. Heart 2014;100:396—405.

[19] Dooley C, Kaur R, Sobieraj DM. Comparison of the efficacy and safety of low molecular weight heparins for venous thromboembolism prophylaxis in medically ill patients. Curr Med Res Opin 2014;30:367—80.

[20] Harenberg J, Marx S, Dahl OE, Marder VJ, Schulze A, Wehling M, et al. Interpretation of endpoints in a network meta-analysis of new oral anticoagulants following total hip or total knee replacement surgery. Thromb Haemost 2012;108:903—12.

[21] Landoni G, Greco T, Biondi-Zoccai G, Nigro Neto C, Febres D, Pintaudi M, et al. Anaesthetic drugs and survival: a Bayesian network meta-analysis of randomized trials in cardiac surgery. Br J Anaesth 2013;111:886—96.

[22] Navarese EP, Buffon A, Andreotti F, Kozinski M, Welton N, Fabiszak T, et al. Meta-analysis of impact of different types and doses of statins on new-onset diabetes mellitus. Am J Cardiol 2013;111: 1123—30.

[23] Phung OJ, Kahn SR, Cook DJ, Murad MH. Dosing frequency of unfractionated heparin thromboprophylaxis: a meta-analysis. Chest 2011;140:374—81.

[24] Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. JAMA 2003;289:2534—44.

[25] Roskell NS, Lip GY, Noack H, Clemens A, Plumb JM. Treatments for stroke prevention in atrial fibrillation: a network meta-analysis and indirect comparisons versus dabigatran etexilate. Thromb Haemost 2010;104:1106—15.

[26] Sciarretta S, Palano F, Tocci G, Baldini R, Volpe M. Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. Arch Intern Med 2011;171:384—94.

[27] Wu HY, Huang JW, Lin HJ, Liao WC, Peng YS, Hung KY, et al. Comparative effectiveness of renin-angiotensin system blockers and other antihypertensive drugs in patients with diabetes: systematic review and bayesian network meta-analysis. BMJ 2013;347:f6008.

[28] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. Value Health 2011;14:429—37.

[29] Higgins JP, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? Lancet 2015;386:628—30.

[30] Mills EJ, Kanters S, Thorlund K, Chaimani A, Veroniki AA, Ioannidis JP. The effects of excluding treatments from network meta-analyses: survey. BMJ 2013;347:f5195.