



OPEN ACCESS

# Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study

Adriani Nikolakopoulou,<sup>1</sup> Dimitris Mavridis,<sup>2,3</sup> Toshi A Furukawa,<sup>4</sup> Andrea Cipriani,<sup>5,6</sup> Andrea C Tricco,<sup>7,8</sup> Sharon E Straus,<sup>7,9</sup> George C M Siontis,<sup>10</sup> Matthias Egger,<sup>1</sup> Georgia Salanti<sup>1</sup>

For numbered affiliations see end of article.

Correspondence to: G Salanti [georgia.salanti@ispm.unibe.ch](mailto:georgia.salanti@ispm.unibe.ch)  
Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;360:k585  
<http://dx.doi.org/10.1136/bmj.k585>

Accepted: 22 January 2018

## ABSTRACT

### OBJECTIVE

To examine whether the continuous updating of networks of prospectively planned randomised controlled trials (RCTs) (“living” network meta-analysis) provides strong evidence against the null hypothesis in comparative effectiveness of medical interventions earlier than the updating of conventional, pairwise meta-analysis.

### DESIGN

Empirical study of the accumulating evidence about the comparative effectiveness of clinical interventions.

### DATA SOURCES

Database of network meta-analyses of RCTs identified through searches of Medline, Embase, and the Cochrane Database of Systematic Reviews until 14 April 2015.

### ELIGIBILITY CRITERIA FOR STUDY SELECTION

Network meta-analyses published after January 2012 that compared at least five treatments and included at least 20 RCTs. Clinical experts were asked to identify in each network the treatment comparison of greatest clinical interest. Comparisons were excluded for which direct and indirect evidence disagreed, based on side, or node, splitting test ( $P < 0.10$ ).

### OUTCOMES AND ANALYSIS

Cumulative pairwise and network meta-analyses were performed for each selected comparison. Monitoring boundaries of statistical significance were constructed and the evidence against the null hypothesis was considered to be strong when the monitoring boundaries were crossed. A significance

level was defined as  $\alpha = 5\%$ , power of 90% ( $\beta = 10\%$ ), and an anticipated treatment effect to detect equal to the final estimate from the network meta-analysis. The frequency and time to strong evidence was compared against the null hypothesis between pairwise and network meta-analyses.

### RESULTS

49 comparisons of interest from 44 networks were included; most ( $n = 39$ , 80%) were between active drugs, mainly from the specialties of cardiology, endocrinology, psychiatry, and rheumatology. 29 comparisons were informed by both direct and indirect evidence (59%), 13 by indirect evidence (27%), and 7 by direct evidence (14%). Both network and pairwise meta-analysis provided strong evidence against the null hypothesis for seven comparisons, but for an additional 10 comparisons only network meta-analysis provided strong evidence against the null hypothesis ( $P = 0.002$ ). The median time to strong evidence against the null hypothesis was 19 years with living network meta-analysis and 23 years with living pairwise meta-analysis (hazard ratio 2.78, 95% confidence interval 1.00 to 7.72,  $P = 0.05$ ). Studies directly comparing the treatments of interest continued to be published for eight comparisons after strong evidence had become evident in network meta-analysis.

### CONCLUSIONS

In comparative effectiveness research, prospectively planned living network meta-analyses produced strong evidence against the null hypothesis more often and earlier than conventional, pairwise meta-analyses.

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Network meta-analysis is an extension to conventional meta-analysis, which includes both direct and indirect evidence on the comparative effectiveness of multiple treatments

Network meta-analysis might produce strong evidence against the null hypothesis on the comparative effectiveness of treatments earlier than standard pairwise meta-analysis but requires more assumptions and advanced statistical methods

Sequential methods for analysis of “living” network meta-analysis of prospectively planned randomised controlled trials have recently become available, allowing the continuous updating of evidence

## WHAT THIS STUDY ADDS

Network meta-analysis was 20% more likely to provide strong evidence against the null hypothesis of treatment differences than pairwise meta-analysis

Network meta-analysis provided strong evidence against the null hypothesis four years earlier than pairwise meta-analysis

Prospectively planned living network meta-analysis can facilitate timely recommendations and contribute to reduce research waste by providing strong evidence against the null hypothesis earlier than living pairwise meta-analysis

## Introduction

A timelier introduction of effective medical interventions was one of the early promises of meta-analysis of randomised control trials (RCTs).<sup>1 2</sup> Cumulative meta-analysis, defined as updating a meta-analysis whenever a new eligible RCT becomes available, has been used retrospectively to examine how evidence on a given intervention has accrued over time and how quickly it has informed guidelines.<sup>3 4</sup> More recently, the optimal time for updating a systematic review has been discussed<sup>5-7</sup> and guidelines and decision tools developed.<sup>8-10</sup> In 2014 “living systematic reviews” were proposed as a framework for continuously updated meta-analyses.<sup>11</sup>

In recent years, network meta-analyses have gained prominence in comparative effectiveness research.<sup>12 13</sup> They extend conventional, pairwise meta-analysis to compare multiple treatments within a network of RCTs.<sup>14-16</sup> A living version of network meta-analysis

has recently been suggested as the new paradigm in comparative effectiveness research.<sup>17–18</sup> Healthcare institutions such as the UK National Institute for Health and Care Excellence and the World Health Organization consider network meta-analyses and, if there is high confidence in the results, use them to inform recommendations.<sup>19</sup> By including both direct and indirect evidence, continuously updated network meta-analysis can reach robust conclusions on the relative effectiveness of treatments earlier than pairwise meta-analyses, thus potentially facilitating timely recommendations and reducing research waste.<sup>17–20</sup>

In a prospectively planned network meta-analysis, studies are designed and realised using a predefined protocol and they are cumulatively synthesised as their results become available. One study highlighted the potential of this approach to optimally inform comparative effectiveness of drugs, not only at the post-marketing stage but also before licensing.<sup>21</sup> In the framework of a prospective living network meta-analysis, suitable methods are required for statistically monitoring the accumulating evidence while controlling for the risk of falsely concluding superiority of an intervention. Such methods have been developed recently, extending the sequential monitoring of trials and pairwise meta-analyses.<sup>18,22</sup> It is, however, unclear whether the theoretical potential of prospectively planned living network meta-analysis can be realised in comparative effectiveness research and whether its increased power compared with pairwise meta-analysis is or is not substantial. We used sequential monitoring to assess recently published network meta-analyses of RCTs of medical interventions to examine whether living network meta-analysis would have provided strong evidence against the null hypothesis more often and earlier than the corresponding updated pairwise meta-analysis.

## Methods

### Search strategy and inclusion criteria

We compiled a large database of network meta-analyses of RCTs based on searches of Medline, Embase, and the Cochrane Database of Systematic Reviews from inception to 14 April 2015.<sup>12</sup> In the present study we included networks published after January 2012, as empirical evidence suggested that the quality of the systematic reviews and statistical rigor have considerably improved in recent years.<sup>12</sup> To ensure a critical mass of data, we included networks with at least one closed loop of evidence that compared at least five different treatments and included 20 or more RCTs published within at least 10 years.

### Selection of comparison of interest

We focused on treatment comparisons that were of interest to the developers of clinical guidelines during the period the body of evidence accumulated. For each included network meta-analysis we asked senior clinicians or clinical researchers with experience in guideline development to choose the treatment comparison that “was of greatest interest to the

developers of guidelines or which had the greatest influence on clinical decision-making during the indicated time period” and to justify their choice by providing a reference of a relevant clinical guideline. One expert evaluated each network and the comparison was chosen independently of the availability of direct or indirect evidence. Experts were blind to the results of the sequential analysis. Treatment effects were expressed as the standardised mean difference, odds ratios, or hazard ratios for continuous, binary, and time-to-event data, respectively.

Network meta-analysis rests on the assumption of consistency between direct and indirect evidence. We therefore excluded networks where the comparison of interest showed evidence of inconsistency, defined by a P value less than 0.10 when direct and indirect evidence were compared in a z test (Separate Indirect from Direct Evidence (SIDE), also called node splitting test).<sup>23</sup>

### Construction of monitoring boundaries and definition of strong evidence

We assumed that studies had been prospectively planned and that they were included in the synthesis model once results became available. Then we evaluated the evidence against the null hypothesis using hypothesis testing to decide whether further data were needed. Repeatedly testing whenever new evidence is added to a body of evidence leads to inflated type I errors.<sup>22–25</sup> Methods originally developed for sequential analysis of RCTs have been adapted to cumulative pairwise meta-analysis and more recently to network meta-analysis.<sup>18,26</sup> We used an adaptation of  $\alpha$  spending functions, which we have described in detail elsewhere.<sup>18</sup> Firstly, we defined an anticipated treatment effect to detect rates of type I and type II errors ( $\alpha$  and  $\beta$ , respectively). Secondly, we constructed the  $\alpha$  spending function boundary as a function of the statistical information (added at each update) and the maximum information. Statistical information is defined as the inverse of the variance—that is, precision. We defined the maximum information as the precision of a single RCT that is adequately powered to detect the anticipated difference between the two interventions, given  $\alpha$  and  $\beta$ . The  $\alpha$  error is then distributed along the sequential tests, with smaller values “spent” for early tests and larger values spent at later stages. The monitoring boundaries correspond to the quantiles of the  $\alpha$  levels and approximate the  $(1-\alpha/2)$  % quantile of the standard normal distribution as the statistical information approaches its maximum. We implemented the methods in a freely available R package (see appendix N).

We defined a significance level  $\alpha=5\%$ , power of 90% ( $\beta=10\%$ ) and an anticipated treatment effect to detect equal to the final estimate from the network meta-analysis. We expressed results as z scores (ie, the effect size divided by its standard error). In the primary analysis for both pairwise meta-analysis and network meta-analysis, we imputed the median value of the empirical distributions from Cochrane reviews.<sup>27,28</sup> In a sensitivity analysis, we estimated heterogeneity

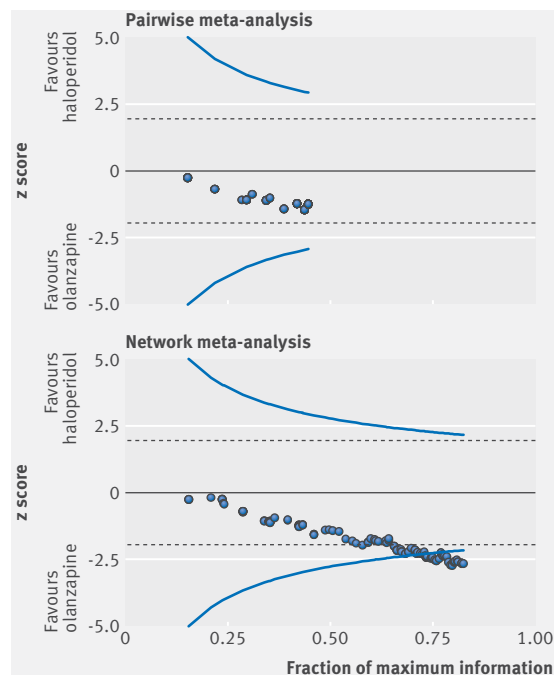
from the data at each update. We considered that a pairwise or network meta-analysis provided strong evidence against the null hypothesis (the hypothesis that there is no difference between the interventions) when the accumulated information crossed the monitoring boundaries of statistical significance, constructed as described here and previously.<sup>18</sup> We define strong evidence as strong evidence against the null hypothesis.

#### Example: olanzapine versus haloperidol in schizophrenia

We illustrate the approach using the example of the relative efficacy of olanzapine and haloperidol in the acute treatment of schizophrenia, based on one of the network meta-analyses included in this study<sup>29</sup> (fig 1). We assume a standardised mean difference measuring the overall change in symptoms of 0.13 favouring olanzapine, equal to the final estimate from the random effects network meta-analysis, and type I and type II errors of 5% and 10%, respectively. The first RCT to compare olanzapine and haloperidol was published in 1996.<sup>30</sup> The results showed that olanzapine tended to reduce symptoms more than haloperidol: the standardised mean difference was 0.07 and the z score was 0.26.<sup>30</sup> Until 2007, a further 10 RCTs that directly compared the two drugs were published, resulting in a summary standardised mean difference of 0.12 favouring olanzapine (95% confidence interval -0.07 to 0.30); these results are added sequentially in cumulative pairwise meta-analysis (fig 1). In network meta-analysis the z score is updated whenever new direct or indirect evidence becomes available (fig 1). Indirect evidence accumulates through RCTs that compare the drugs of interest with placebo or another drug. At any time, the accumulated information is compared with the monitoring boundaries. The direct evidence from cumulative pairwise meta-analysis remains within the boundaries, whereas the mixed evidence (direct and indirect) from network meta-analysis crosses the monitoring boundaries in 2008 (after the inclusion of 131 RCTs in the entire network), indicating strong evidence against the null hypothesis for the superiority of olanzapine (fig 1). The standardised mean difference at the point of crossing the stopping boundary was 0.13 (95% confidence interval 0.02 to 0.24). Appendix M shows an alternative presentation of sequential monitoring using repeated confidence intervals.

#### Comparing living network meta-analysis and living pairwise meta-analysis

We compared the emergence of strong evidence against the null hypothesis (defined as crossing monitoring boundaries) between living pairwise and network meta-analyses. Among comparisons with strong evidence we examined whether boundaries were crossed as a result of indirect evidence. We analysed data in a 2×2 table using McNemar's exact test and estimated differences in the probability of providing strong evidence. We plotted Kaplan-Meier curves and



**Fig 1 | Efficacy of olanzapine versus haloperidol in treatment of acute schizophrenia, as estimated from living pairwise meta-analysis and living network meta-analysis. Monitoring boundaries were constructed using an  $\alpha$  spending function with type I and type II errors fixed at 5% and 10%, respectively. Conventional significance thresholds are shown as dotted lines ( $z=1.96$ ). The horizontal axis shows statistical information that accumulated over time, compared with maximum statistical information (information in single adequately powered study). Heterogeneity variance is assumed to be equal to the median of predictive distributions (0.049)**

calculated the hazard ratio from a frailty time-to-event regression model to describe the time needed to cross the boundary while accounting for the paired nature of the data.<sup>31</sup> For comparisons with strong evidence against the null hypothesis we recorded how many studies that directly compared the treatments were published after the boundaries had been crossed. All analyses were done in a frequentist framework using R (R Development Core Team, Vienna, Austria) and Stata (Stata, College Station, TX). Appendices K and N include the technical details about the R package, and worked examples.

#### Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in the design and implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

#### Results

##### Database of network meta-analyses

Out of 456 published network meta-analyses included in the original database,<sup>12</sup> 44 met the inclusion

criteria. The most important reasons for exclusion were publication before January 2012 and lack of outcome data (see appendix figure 1). The 44 network meta-analyses were published in 38 journals (28 specialist and 10 general medicine journals) (table 1). Most networks addressed research questions in the specialties of cardiology, endocrinology, psychiatry, and rheumatology (table 1). Clinical experts selected 54 treatment comparisons, and five were excluded owing to evidence of inconsistency (table 2, and see appendix table 1). Most of the 49 included comparisons were between two drug interventions (n=39, 80%). Five comparisons (10%) involved placebo, two (4%) were between invasive interventions, and four (8%) involved

lifestyle modifications. Most primary outcomes were binary (66%) or continuous (23%) (table 1).

Of the 49 comparisons, 29 (59%) were informed by both direct and indirect evidence in the network meta-analysis. The P values from testing for agreement between indirect and direct evidence ranged between 0.11 and 0.99 (see appendix table 1). Thirteen comparisons (27%) were not examined directly in any RCT; seven (14%) were based on direct evidence only.

### Comparison of living pairwise and network meta-analyses

For 10 of the 49 comparisons (20%), the evidence for superiority of one of the interventions was stronger with network meta-analysis than with pairwise meta-analysis. In seven instances (14%) both pairwise and network meta-analyses provided strong evidence against the null hypothesis, whereas in 32 comparisons neither analysis produced strong evidence (table 3, P=0.002 from McNemar's exact test). Network meta-analysis was 20% more likely (95% confidence interval 10% to 35%) to provide strong evidence against the null hypothesis than pairwise meta-analysis. Results were similar when heterogeneity was estimated rather than imputed (see appendix table 2) or when summary effects from pairwise meta-analysis instead of network-meta-analysis were used to define the anticipated treatment effect to detect (see appendix table 3). Restricting analyses to comparisons for which both direct and indirect evidence were available did not materially change results (P=0.016 from McNemar's test, see appendix table 4).

For nine out of the 17 treatment comparisons where network meta-analyses provided strong evidence against the null hypothesis (53%), this was achieved only after adding an RCT that contributed indirect evidence. For 13 treatment comparisons there was no RCT directly comparing the interventions of interest, and yet for three of them strong evidence was available by indirect comparison (table 2): exercise and non-steroidal anti-inflammatory drugs versus exercise for pain relief,<sup>32</sup> nifedipine versus placebo for delaying delivery in women at risk of preterm delivery,<sup>33</sup> and candesartan versus topiramate for the prevention of migraine.<sup>34</sup>

Median time to strong evidence against the null hypothesis (the first time that the monitoring boundary was crossed) was 19 years (interquartile range 16 to 23) with network meta-analysis and 23 years (interquartile range not estimable) with pairwise meta-analysis (fig 2). Network meta-analysis provided strong evidence earlier than pairwise meta-analysis, by 4 years (95% confidence interval 0 to 7 years). The hazard ratio comparing network with pairwise meta-analysis was 2.78 (95% confidence interval 1.00 to 7.72; P=0.05). For eight (47%) of the 17 comparisons with strong evidence, studies directly comparing the treatments of interest continued to be published after the boundary had been crossed (see appendix table 1). The total number of additional studies was 66; 40 of these compared edaravone with placebo.<sup>35</sup> Appendix

**Table 1 | Characteristics of 44 network meta-analyses and 49 network comparisons of medical interventions included in study. Values are numbers (percentages) unless stated otherwise**

Network meta-analyses	Estimates
Median (interquartile range):	n=44
No of studies included	41 (26-60)
No of treatments compared	9 (8-14)
Span of years	21 (16-34)
Total sample size	10 587 (3494-26 089)
Outcome characteristics	
Direction of effect:	
Beneficial	22 (50)
Harmful	22 (50)
Outcome type:	
Objective	17 (39)
Semi-objective*	19 (43)
Subjective†	8 (18)
Measurement:	
Binary	29 (66)
Continuous	10 (23)
Rate	1 (2)
Time to event	4 (9)
Journal (No of distinct journals):	
General medicine (n=10)	16 (36)
Specialty (n=28)	28 (64)
Medical specialty:	
Cardiology	9 (20)
Endocrinology	6 (13)
Psychiatry	5 (11)
Rheumatology	5 (11)
Neurology	3 (7)
Dentistry/periodontology	3 (7)
Pulmonology	3 (7)
Dermatology	2 (5)
Gastroenterology	2 (5)
Obstetrics	2 (5)
Oncology	2 (5)
Anaesthesiology	1 (2)
Hepatology	1 (2)
Comparisons:	
n=49	
Drug versus drug	39 (80)
Drug versus placebo	4 (8)
Lifestyle versus drug	2 (4)
Lifestyle versus lifestyle and placebo	1 (2)
Lifestyle versus lifestyle and drug	1 (2)
Invasive versus invasive	2 (4)

\*Cause specific mortality, major morbidity event, composite mortality or morbidity, obstetric outcomes, internal structure, external structure, surgical device success or failure, withdrawals or drop-outs, resource use, and hospital stay or process measures.<sup>27</sup>

†Pain, mental health outcomes, dichotomous biological markers, quality of life or functioning, consumption, satisfaction with care, general physical health, adverse events, infection or new disease, continuation or termination of condition being treated, and composite endpoint (including at most one mortality or morbidity endpoint).<sup>27</sup>

**Table 2 | Treatment comparisons selected in each network, type of evidence (direct, indirect, or both), and meta-analysis method that provides strong evidence against similarity of treatments for primary outcome studied (see appendix for more detailed version of table)**

Network	Comparison of greatest interest	Type of evidence	Design providing strong evidence	
			Network meta-analysis	Meta-analysis
Buti 2013	Coronally advanced flap and connective tissue graft versus coronally advanced flap and enamel matrix derivative	Both	No	No
Dogliotti 2013	Vitamin K antagonists versus apixaban	Both	No	No
Naci 2013	Atorvastatin versus rosuvastatin	Both	No	No
Filippini 2013	$\beta$ interferon-1a (Avonex) versus $\beta$ interferon-1a (Rebif)	Both	No	No
Hon-Yen Wu 2013	Angiotensin receptor blockers versus angiotensin converting enzyme inhibitors	Both	No	No
Lin 2014	Ferric sulphate versus mineral trioxide aggregate	Both	No	No
Castellucci 2014	Unfractionated heparin and vitamin K antagonist versus low molecular weight heparin and vitamin K antagonist	Both	Yes	Yes
Myers 2014	Celecoxib versus tramadol	Both	No	No
Alfirevic 2014	Vaginal misoprostol versus vaginal prostaglandin E2	Both	No	No
Greco 2015	Dobutamine versus levosimendan	Both	Yes	No
Greco 2015	Dobutamine versus milrinone	Both	No	No
Walsem 2015	Diclofenac high dose versus celecoxib	Both	Yes	No
Singh 2015	5-aminosalicylic acid versus anti-tumour necrosis factor	Both	Yes	No
Singh 2015	Anti-tumour necrosis factor versus placebo	Both	Yes	No
Linde 2015	Selective serotonin reuptake inhibitors versus tricyclic and tetracyclic antidepressants	Both	No	No
Sun 2015	Metformin versus sitagliptin	Both	Yes	No
Leucht 2013	Haloperidol versus olanzapine	Both	Yes	No
Ke-Qing Shi 2013	Endoscopic injection sclerotherapy versus endoscopic banding ligation	Both	Yes	No
Stagg 2014	Isoniazid (six months) versus rifampicin and isoniazid (three-four months)	Both	No	No
Tadrous 2014	Alendronate versus risedronate	Both	No	No
Dong 2013	Inhaled corticosteroids versus long acting $\beta$ 2 agonists—inhaled corticosteroids	Both	No	No
Stevens 2015	Standard care or placebo versus diet and exercise	Both	Yes	Yes
Lin 2012	Chemical occlusion versus physical occlusion	Both	No	No
Fretheim 2012	Angiotensin converting enzyme inhibitors versus calcium channel blockers	Both	No	No
Fretheim 2012	Angiotensin receptor blockers versus calcium channel blockers	Both	No	No
Liu 2012	Thiazolidinediones versus dipeptidyl peptidase-4 inhibitors	Both	No	No
Lori 2012	Amiodarone intravenous versus flecainide intravenous	Both	No	No
Ara 2012	Orlistat versus standard care	Both	Yes	Yes
Gray 2012	Orlistat versus lifestyle	Both	Yes	Yes
Chatterjee 2013	Metoprolol versus bisoprolol	Indirect	No	NE
Mavranouzouli 2013	Sertraline versus diazepam	Indirect	No	NE
Akshintala 2013	Non-steroidal anti-inflammatory drugs versus nafamostat	Indirect	No	NE
Bodalia 2013	Pregabalin versus gabapentin	Indirect	No	NE
Kew 2014	Glycopyrronium bromide versus budesonide	Indirect	No	NE
Windecker 2014	Everolimus eluting stent versus coronary artery bypass grafting	Indirect	No	NE
Kriston 2014	Fluoxetine versus escitalopram	Indirect	No	NE
Dong 2015	Exercise and non-steroidal anti-inflammatory drugs versus exercise	Indirect	Yes	NE
Rotta 2013	Terbinafine versus flutrimazole	Indirect	No	NE
Murad 2012	Alendronate versus denosumab	Indirect	No	NE
Ramsberg 2012	Amitriptyline versus fluoxetine	Indirect	No	NE
Haas 2012	Nifedipine versus placebo	Indirect	Yes	NE
Shamiliyan 2012	Candesartom versus topiramate	Indirect	Yes	NE
Shi 2013	Endoscopic banding ligation versus endoscopic banding ligation and endoscopic injection sclerotherapy	Direct	No	No
Dogliotti 2013	Vitamin K antagonists versus rivaroxab	Direct	No	No
Pechlivanoglou 2013	Posaconazole versus fluconazole	Direct	Yes	Yes
Yang 2014	Edaravone versus placebo	Direct	Yes	Yes
Zoccai 2014	Iopromide versus iodixanol	Direct	No	No
Samarasekera 2013	Potent corticosteroid versus placebo	Direct	Yes	Yes
Terasawa 2012	Fludarabine-rituximab-based chemoimmunotherapies versus fludarabine based combination regimens	Direct	No	No

NE=not estimable.

table 5 shows the results from pairwise and network meta-analyses for each medical specialty.

In a scatterplot of the precision of estimates from pairwise and network meta-analyses, there was a clear gain in precision with network meta-analysis (see appendix figure 2). In almost half of the comparisons (13 out of 29 comparisons, 45%) the

network meta-analysis produced a 95% confidence interval the width of which was less than two thirds of the interval from pairwise meta-analysis. Appendix figure 3 presents the continuous updating (as z scores along with monitoring boundaries) of pairwise and network meta-analysis for all 49 comparisons included in this study.

**Table 3 | Number of comparisons with strong evidence against the null hypothesis from pairwise and network meta-analysis. Values are numbers (percentages) unless stated otherwise**

Pairwise meta-analysis	Network meta-analysis		Total	P value*
	Yes	No		
Yes	7 (14)	0 (0)	7 (14)	0.002
No	10 (20)	32 (65)	42 (86)	
<b>Total</b>	17 (35)	32 (65)	49 (100)	

\*McNemar's exact test.

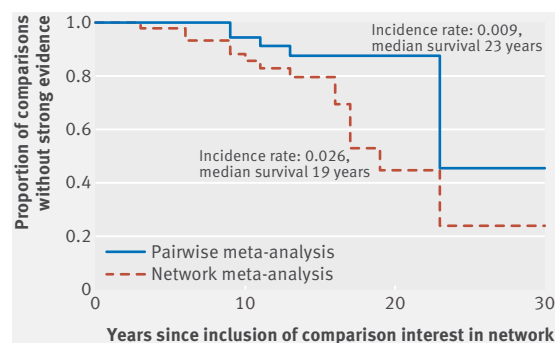
The anticipated relative treatment effect was set equal to the final estimate from network meta-analysis. Heterogeneity variance was imputed as the median value of its empirical distribution.<sup>27 28</sup>

## Discussion

This study found that among 49 treatment comparisons deemed important for guideline development and clinical practice, prospectively planned, living network meta-analysis was 20% more likely to produce strong evidence against the null hypothesis than living pairwise meta-analysis that was based on direct evidence only. Strong evidence became available four years earlier with network meta-analysis compared with pairwise meta-analysis. Of note, studies comparing the two treatments of interest continued to be published even after strong evidence had become available. This is an important finding with implications for clinical research, especially in the context of the debate about research waste.<sup>20 36-38</sup> As per the inclusion criteria, the findings of this study apply to treatment comparisons for which a considerable amount of data have accumulated over at least 10 years.

## Strengths and weaknesses of this study

Several authors have argued that network meta-analyses and indirect treatment comparisons should be more frequently used to inform healthcare and regulatory decisions.<sup>39-43</sup> One of these studies analysed a network of interventions for primary open angle glaucoma and found that network meta-analysis showed the advantage of prostaglandins 10 years before they were recommended in clinical guidelines.<sup>39</sup> In our study we empirically assessed the frequency of and time to



**Fig 2 | Kaplan-Meier survival curves for non-strong evidence against null hypothesis, comparing sequential pairwise and network meta-analysis of 49 comparisons. Events occur when monitoring boundaries are crossed for comparison of interest. Time is measured as years from time point when both interventions are included in network**

strong evidence in comparative effectiveness research using network or traditional pairwise meta-analysis. Although the sample size was small (49 comparisons) we believe it is likely to represent situations where guideline developers and clinical decision makers might consider network meta-analysis.

We mimicked the situation of a prospectively planned living network or pairwise meta-analysis and asked clinical experts to choose the treatment comparisons that were of topical interest during the period the evidence accumulated. We restricted the data to networks that did not show evidence of statistical inconsistency for the comparisons of interest, and excluded only five networks with evidence of inconsistency; this is in line with previous studies that have shown that direct and indirect comparisons in a network disagree in about 10% of comparisons.<sup>44 45</sup> However, we cannot rule out inconsistency in some of the comparisons evaluated; statistical tests for inconsistency have low power to detect inconsistency.<sup>46 47</sup> One in eight networks was previously found to show evidence of inconsistency using the design-by-treatment test; this means that our methods would not be applicable in, on average, one in eight networks.<sup>44</sup>

Our study has some limitations. We reanalysed published networks that did not show statistical inconsistency, but we did not examine the overall quality of the evidence they provided. We acknowledge that strong evidence against the null hypothesis does not necessarily translate into strong recommendations.<sup>48</sup> Guideline panels typically consider the quality of evidence in the results from meta-analysis before making recommendations, often using the GRADE system.<sup>49</sup> Evaluation of the confidence in the results from network meta-analysis is a matter of ongoing research,<sup>50-52</sup> and none of the included networks in our database attempted such an evaluation. We did not consider other components such as the risk of publication bias, the limitations of the RCTs included in the networks, or the comprehensiveness of the literature search and accuracy of the data extraction. Selective publication of studies of the comparison of interest or of studies of comparisons that contributed indirect evidence or a high risk of bias in the conduct and analysis of studies, could have affected our conclusions. Finally, we acknowledge that evidence against the null hypothesis based on P values might be of greater interest to regulatory agencies than to guideline developers. Interpretation of the 95% confidence interval of the treatment effect in the context of worthwhile effects are more useful for decision making.<sup>53 54</sup> Although recently published networks, such as those included here, conform with high methodological standards,<sup>55</sup> the strength of recommendations from pairwise and network meta-analyses should be compared in future studies.

## Relation to other studies and implications

Protocols for living pairwise and network meta-analyses have been published recently,<sup>56-58</sup> and

health technology assessment agencies, the World Health Organization, and drug licensing bodies have recognised the value of synthesising direct and indirect evidence in network meta-analysis.<sup>19 21 41 59</sup> The concept of living meta-analysis is in line with the goal to continuously update guidelines and to promptly translate results to recommendations, evidence summaries, and decision aids.<sup>60 61</sup> Our findings suggest that the gain in including both direct and indirect evidence when synthesising data from clinical research is substantial as it can considerably reduce the time to strong evidence against the null hypothesis.

We found that one in four of the comparisons was not directly compared in RCTs. This might be partly explained by the tendency of pharmaceutical companies to test drugs against placebo or suboptimal interventions, rather than the reference treatment given in routine practice.<sup>62</sup> Taking into account direct and indirect evidence, the median time to obtain strong evidence against the null hypothesis about comparisons of interest was 19 years. Consequently, even in cases where direct evidence exists, it is

important to strengthen the evidence base using network meta-analysis.

#### Unanswered questions and future work

Our results indicate that living network meta-analysis has the potential to reduce research waste and optimise the use of available evidence.<sup>17</sup> Further research is needed to better understand the role of living meta-analysis in the prevention of research waste. Previous studies showed that the impact of available data on the design of subsequent research is generally low.<sup>63 66</sup> We found that in about half of the comparisons for which strong evidence emerged from living network meta-analysis, further studies were conducted after the evidence was already available. We did not assess these studies to determine whether they were scientifically and ethically justified but are planning such work in the future.

Methodologists are debating the use of sequential methods for cumulative meta-analysis. In particular, there is concern about encouraging inference based on statistical rather than clinical significance (see box). The statistical routines are not widely available, and

#### Box 1: Do researchers agree on the use of sequential methods in meta-analysis?

Consensus on the appropriateness of applying sequential methods in meta-analysis is lacking. The arguments against their use are mainly twofold: the suspicion that sequential methods encourage inferences on the basis of statistical significance and concerns about meta-analysis influencing decisions on the design of future trials.<sup>1 2</sup> In our view, interpretation of meta-analysis results should emphasise the uncertainty surrounding effect estimates, irrespective of the use of sequential analysis. Uncertainty over stopping decisions can be expressed and inspected through repeated confidence intervals, which can be drawn in a forest plot along with confidence intervals (see appendix M)<sup>3 4</sup>

Concerns about the influence of meta-analyses on future primary research has also contributed to the scepticism. In particular, detractors of the notion of sequential meta-analysis have questioned the analogy between stopping trials and stopping meta-analyses and wondered whether a decision of “no further updating warranted” would be reasonable.<sup>2</sup> Editors of Cochrane review groups have indeed “closed a review,” judging that it is not likely that further evidence would challenge current conclusions, and therefore new trials are deemed unnecessary, costly, and unethical.<sup>5</sup> Authors of systematic reviews are not in a position to decide whether further studies are done and to define their design. What they should do, however, is to provide recommendations for research, to identify potential gaps in the existing evidence base, and to discuss the implications of their reviews for future research.<sup>2 3 7-11</sup>

The application of sequential methods in meta-analysis is also controversial; this controversy is mainly driven by the nature of their use and interpretation in practice. Sequential methods are often used to correct for multiple testing when presenting and interpreting evidence from a systematic review. In our view, no multiplicity is induced in the conventional process of performing meta-analysis as a retrospective activity: researchers simply synthesise what is already known and are not in a position to decide on carrying out further studies.<sup>2</sup> Retrospective application of sequential methods in meta-analysis in line with recommendations for cumulative meta-analysis<sup>7</sup> should be done for illustrative purposes only. However, in a prospective meta-analysis setting where studies are planned and analysed sequentially to answer a research question, control of type I error through sequential methods is desirable.<sup>12</sup> Researchers undertaking such living systematic reviews have to decide a priori and describe in a protocol if and when the review is going to be terminated. If this decision is linked to whether treatment effects provide strong evidence against the null hypothesis or not, adjustment for multiple testing becomes important. Empirical results presented in this paper should be seen as an illustration of hypothetical living networks of prospectively planned studies rather than as an attempt to define the need of future research in the examined healthcare areas. Depending on whether researchers plan to simply describe implications for research, provide concrete recommendations for filling evidence gaps, or actively direct future research, use of sequential methods might be less or more imperative. In our view, the optimal use of living systematic reviews will be to highlight gaps and certainties in a body of evidence and provide research funders and regulators with the best evidence to decide whether new primary research is warranted and, if so, for which treatment comparisons.

Undertaking living systematic reviews within a bayesian framework is a possible alternative.<sup>3 13</sup> The estimated treatment effects after each update form a prior for future updates. Then, approaches to monitor the accumulated evidence can be informal (eg, inspecting the estimated treatment effects and their precision without formally specifying a stopping criterion) or formal (eg, by defining a loss function or a boundary as a basis for monitoring).<sup>14 15</sup>

1. Chalmers TC, Lau J. Meta-analytic stimulus for changes in clinical trials. *Stat Methods Med Res* 1993;22:161-72.

2. Higgins JPT. Comment on “Trial sequential analysis: methods and software for cumulative meta-analyses” by Wetterslev and colleagues. *Cochrane Methods. Cochrane DB Syst Rev* 2012;(Suppl 1):1-56. [www.cochranelibrary.com/dotAsset/3d4dc937-0b49-4634-a766-ef13df169f9f.pdf](http://www.cochranelibrary.com/dotAsset/3d4dc937-0b49-4634-a766-ef13df169f9f.pdf).

3. Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med* 2011;309:903-21.

4. Jennison C, Turnbull BW. Repeated confidence intervals for group sequential clinical trials. *Control Clin Trials* 1984;5:33-45.

5. Lacasse Y, Cates CJ, McCarthy B, Welsh EJ. This Cochrane Review is closed: deciding what constitutes enough research and where next for pulmonary rehabilitation in COPD. <http://doi.wiley.com/10.1002/14651858.ED000107>

(Continued)

## Box 1: (Continued)

6. Sutton AJ, Donegan S, Takwoingi Y, Garner P, Gamble C, Donald A. An encouraging assessment of methods to inform priorities for updating systematic reviews. *J Clin Epidemiol* 2009;623:241-51.
7. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Wiley; 2011:434.
8. Chapman E, Reveiz L, Chambliss A, Sangalang S, Bonfill X. Cochrane systematic reviews are useful to map research gaps for decreasing maternal mortality. *J Clin Epidemiol* 2013;661:105-12.
9. Higgins JP, Green S, Scholten RJ. Maintaining Reviews: Updates, Amendments and Feedback. In: Higgins JP, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: Wiley; 2008:31-49 [cited 2014 Jul 16]. <http://doi.wiley.com/10.1002/9780470712184.ch3>
10. Habre C, Tramèr MR, Pöpping DM, Elia N. Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection. *BMJ* 2014;348:g5219.
11. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med* 2014;18:112:e1001603.
12. Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Stat Med* 1997;1624:2901-13.
13. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. *Stat Med* 2016; 1 Aug.
14. Spiegelhalter DJ, Abrams KR, Myles JP. Randomised controlled trials. In: Bayesian approaches to clinical trials and health-care evaluation. Wiley; 2003:181-249.
15. Freedman LS, Spiegelhalter DJ. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Control Clin Trials* 1989;104:357-67.

guidance and tutorials are lacking. The development of software tools and a template protocol for living systematic reviews are urgently needed. Further development of the methodology, such as the most appropriate approach to dealing with heterogeneity in a sequential meta-analysis, is another priority.<sup>11 22</sup>

In the present work, we focused on detecting differences between interventions. It is also possible to construct futility stopping boundaries,<sup>18</sup> and empirical evidence about the relative advantage of network meta-analysis in this context is required. Such an extension might be particularly useful considering that we could not detect statistically significant differences in two out of three comparisons. We selected only one treatment comparison for each network, although decision making may involve several treatments included in the network. Future studies should investigate the superiority or non-inferiority of several competing treatments.

### Conclusions

Continuously updated systematic reviews to inform guidelines and clinical decision making may provide strong evidence against the null hypothesis more frequently and earlier if both direct and indirect accumulating evidence is considered within the framework of a living network meta-analysis.

### AUTHOR AFFILIATIONS

<sup>1</sup>Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland

<sup>2</sup>Department of Primary Education, University of Ioannina, Ioannina, Greece

<sup>3</sup>Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité, Inserm/Université Paris Descartes, Paris, France

<sup>4</sup>Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

<sup>5</sup>Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK

<sup>6</sup>Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK

<sup>7</sup>Knowledge Translation Program, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, Canada

<sup>8</sup>Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

<sup>9</sup>Department of Medicine, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>10</sup>Department of Cardiology, Bern University Hospital, Bern, Switzerland

We thank the following clinical experts who identified clinically relevant comparisons of highest interest and provided a relevant reference to guidelines (numbers in brackets refer to the relevant references to the included network meta-analyses, which can be found in appendix C): Graziella Filippini for reviewing a network on multiple sclerosis,[4] Maria Kyrgiou for reviewing a network on labour induction,[10] Niklaus Meier for reviewing a network on migraine,[44] Nikolaos Pandis for reviewing three networks on dentistry and periodontology,[1, 7, 34] and Stephan Reichenbach for reviewing three networks on rheumatology.[9, 12, 41] We also thank Akhil Parashar, Karan Sud, Apostolos Tsapas, Sara Mazzucco, and Klaus Linde for discussions on the process to select the comparison of highest interest in a previous version of this project; and Maria Petropoulou, Areti-Angeliki Veroniki, Patricia Rios, Afshin Vafaie, Wasifa Zarin, Myrsini Giannatsi, Shannon Sullivan, and Anna Chaimani for their contribution to building the database of network meta-analyses that we used for this study.

**Contributors:** GS and ME conceived and designed the study. TAF, AC, ACT, SES, and GCMS assisted in the design of the study, indicated the relevant comparisons of highest interest for their specialties of expertise, and provided a relevant reference to guidelines. AN compiled the database of eligible networks and performed the analysis. AN drafted the manuscript. TAF and AC assessed network meta-analyses pertaining to mental health, ACT and SES assessed networks in the specialties of endocrinology, dermatology, gastroenterology, obstetrics, oncology, anaesthesiology, and hepatology, and GCMS assessed network meta-analyses from the specialty of cardiology. For networks examining other conditions, external medical doctors or experienced researchers were approached. All authors critically revised the manuscript, interpreted the results, and performed a critical review of the manuscript for intellectual content. GS and ME produced the final version of the submitted article, and all coauthors approved it. AN and GS had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the analysis. GS and ME are the guarantors.

**Funding:** ACT is funded by a tier 2 Canada research chair in knowledge synthesis. SES is funded by a tier 1 Canada research chair in knowledge translation. GS received funding from a Horizon 2020 Marie-Curie individual fellowship (grant No 703254). AC is supported by the National Institute for Health Research Oxford cognitive health clinical research facility. The sponsors had no influence on the design, analysis, and reporting of this study, neither on the preparation of the manuscript.

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare:

no support from any organisation for the submitted work; TAF has received lecture fees from Eli Lilly, Janssen, Meiji, Mitsubishi-Tanabe, MSD, and Pfizer and consultancy fees from Takeda Science Foundation. He has received research support from Mochida and Mitsubishi-Tanabe; no other relationships or activities that could appear to have influenced the submitted work.

**Ethical approval:** Not required.

**Data sharing:** List of included studies and details about the data are reported in the appendix. Study level data and R codes are available in a GitHub repository. Instructions to access them can be found in appendix N.

**Transparency:** The lead author (GS) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ* 1997;315:1371-4. 10.1136/bmj.315.7119.1371
- 2 Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-6. 10.7326/0003-4819-127-9-199711010-00008
- 3 Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-8. 10.1001/jama.1992.03490020088036
- 4 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992;327:248-54. 10.1056/NEJM199207233270406
- 5 Moher D, Tsertsvadze A, Tricco A. When and how to update systematic reviews. In: The Cochrane Collaboration, editor. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2008 [cited 2014 Mar 21]. <http://doi.wiley.com/10.1002/14651858.MR000023.pub3>
- 6 Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007;147:224-33. 10.7326/0003-4819-147-4-200708210-00179
- 7 French SD, McDonald S, McKenzie JE, Green SE. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC Med Res Methodol* 2005;5:33. 10.1186/1471-2288-5-33
- 8 Ahmadzai N, Newberry SJ, Maglione MA, et al. A surveillance system to assess the need for updating systematic reviews. *Syst Rev* 2013;2:104. 10.1186/2046-4053-2-104
- 9 Garner P, Hopewell S, Chandler J, et al. Panel for updating guidance for systematic reviews (PUGs). When and how to update systematic reviews: consensus and checklist. *BMJ* 2016;354:i3507. 10.1136/bmj.i3507
- 10 Takwoingi Y, Hopewell S, Tovey D, Sutton AJ. A multicomponent decision tool for prioritising the updating of systematic reviews. *BMJ* 2013;347:f7191. 10.1136/bmj.f7191
- 11 Elliott JH, Turner T, Clavisi O, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med* 2014;11:e1001603. 10.1371/journal.pmed.1001603
- 12 Petropoulou M, Nikolakopoulou A, Veroniki A-A, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol* 2016;82:20-28.
- 13 Bafeta A, Trinquart L, Seror R, Ravaud P. Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review. *BMJ* 2013;347:f3675.
- 14 Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897-900.
- 15 Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159:130-7. 10.7326/0003-4819-159-2-201307160-00008
- 16 Li T, Puhon MA, Vedula SS, Singh S, Dickersin K, Ad Hoc Network Meta-analysis Methods Meeting Working Group. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med* 2011;9:79. 10.1186/1741-7015-9-79
- 17 Vandvik PO, Brignardello-Petersen R, Guyatt GH. Living cumulative network meta-analysis to reduce waste in research: A paradigmatic shift for systematic reviews? *BMC Med* 2016 Dec [cited 2017 Apr 5];14(1). <http://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0596-4>
- 18 Nikolakopoulou A, Mavridis D, Egger M, Salanti G. Continuously updated network meta-analysis and statistical monitoring for timely decision-making. *Stat Methods Med Res* 2016;962280216659896.
- 19 Kanter 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. 10.2471/BLT.16.174326
- 20 Créquit P, Trinquart L, Yavchitz A, Ravaud P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. *BMC Med* 2016;14:8. 10.1186/s12916-016-0555-0
- 21 Naci H, O'Connor AB. Assessing comparative effectiveness of new drugs before approval using prospective network meta-analyses. *J Clin Epidemiol* 2013;66:812-6. 10.1016/j.jclinepi.2013.04.002
- 22 Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med* 2011;30:903-21. 10.1002/sim.4088
- 23 Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;29:932-44. 10.1002/sim.3767
- 24 Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008;61:64-75. 10.1016/j.jclinepi.2007.03.013
- 25 Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ Open* 2016;6:e011890. 10.1136/bmjopen-2016-011890
- 26 Simmonds M, Salanti G, McKenzie J, Elliott J, Living Systematic Review Network. Living systematic reviews: 3. Statistical methods for updating meta-analyses. *J Clin Epidemiol* 2017;91:38-46. 10.1016/j.jclinepi.2017.08.008
- 27 Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol* 2012;41:818-27. 10.1093/ije/dys041
- 28 Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol* 2015;68:52-60. 10.1016/j.jclinepi.2014.08.012
- 29 Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013;382:951-62. 10.1016/S0140-6736(13)60733-3
- 30 Beasley CMJr, Tollefson G, Tran P, Satterlee W, Sanger T, Hamilton S. Olanzapine versus placebo and haloperidol: acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology* 1996;14:111-23. 10.1016/0893-133X(95)00069-P
- 31 Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979;16:439-54. 10.2307/2061224
- 32 Dong W, Goost H, Lin X-B, et al. Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)* 2015;94:e510. 10.1097/MD.0000000000000510
- 33 Haas DM, Caldwell DM, Kirkpatrick P, McIntosh JJ, Welton NJ. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. *BMJ* 2012;345:e6226. 10.1136/bmj.e6226
- 34 Shamlivan TA, Kane RL, Taylor FR. Migraine in Adults: Preventive Pharmacologic Treatments [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 [cited 2016 Dec 22]. (AHRQ Comparative Effectiveness Reviews). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK138287/>
- 35 Yang B, Shi J, Chen X, Ma B, Sun H. Efficacy and safety of therapies for acute ischemic stroke in China: a network meta-analysis of 13289 patients from 145 randomized controlled trials. *PLoS One* 2014;9:e88440. 10.1371/journal.pone.0088440
- 36 Moher D, Glasziou P, Chalmers I, et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet* 2016;387:1573-86. 10.1016/S0140-6736(15)00307-4
- 37 Storz-Pfennig P. Potentially unnecessary and wasteful clinical trial research detected in cumulative meta-epidemiological and trial sequential analysis. *J Clin Epidemiol* 2017;82:61-70. 10.1016/j.jclinepi.2016.11.003
- 38 Li T, Vedula SS, Scherer R, Dickersin K. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. *Ann Intern Med* 2012;156:367-77. 10.7326/0003-4819-156-5-201203060-00009
- 39 Rouse B, Cipriani A, Shi Q, Coleman AL, Dickersin K, Li T. Network Meta-analysis for Clinical Practice Guidelines: A Case Study on First-Line Medical Therapies for Primary Open-Angle Glaucoma. *Ann Intern Med* 2016;164:674-82. 10.7326/M15-2367

- 40 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. 10.1007/s00406-016-0715-4
- 41 Eichler H-G, Thomson A, Eichler I, Schneeweiss S. Assessing the relative efficacy of new drugs: an emerging opportunity. *Nat Rev Drug Discov* 2015;14:443-4. 10.1038/nrd4664
- 42 Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet* 2015;386:628-30. 10.1016/S0140-6736(15)61478-7
- 43 Naci H, van Valkenhoef G, Higgins JPT, Fleurence R, Ades AE. Evidence-based prescribing: combining network meta-analysis with multicriteria decision analysis to choose among multiple drugs. *Circ Cardiovasc Qual Outcomes* 2014;7:787-92. 10.1161/CIRCOUTCOMES.114.000825
- 44 Veroniki AA, Vasilidis HS, Higgins JPT, Salanti G. Evaluation of inconsistency in networks of interventions. *Int J Epidemiol* 2013;42:332-45. 10.1093/ije/dys222
- 45 Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ* 2011;343:d4909.
- 46 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 Dec [cited 2017 Apr 6];12(1). <http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-138>
- 47 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 Dec [cited 2017 Apr 6];14(1). <http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-106>
- 48 Guyatt GH, Oxman AD, Vist GE, et al. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6. 10.1136/bmj.39489.470347.AD
- 49 Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383-94. 10.1016/j.jclinepi.2010.04.026
- 50 Puhan MA, Schünemann HJ, Murad MH, et al. GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630. 10.1136/bmj.g5630
- 51 Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9:e99682. 10.1371/journal.pone.0099682
- 52 Brignardello-Petersen R, Bonner A, Alexander PE, et al. GRADE Working Group. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol* 2018;93:36-44. 10.1016/j.jclinepi.2017.10.005
- 53 Schünemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol* 2016;75(Supplement C):6-15. 10.1016/j.jclinepi.2016.03.018
- 54 Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials* 1997;18:580-93, discussion 661-6. 10.1016/S0197-2456(97)00051-2
- 55 Zarin W, Veroniki AA, Nincic V, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med* 2017;15:3. 10.1186/s12916-016-0764-6
- 56 Badgett RG, Vindhyal M, Stirnaman JT, Gibson CM, Halaby R. A Living Systematic Review of Nebulized Hypertonic Saline for Acute Bronchiolitis in Infants. *JAMA Pediatr* 2015;169:788-9. 10.1001/jamapediatrics.2015.0681
- 57 Rahal AK, Badgett RG, Hoffman RM. Screening Coverage Needed to Reduce Mortality from Prostate Cancer: A Living Systematic Review. *PLoS One* 2016;11:e0153417. 10.1371/journal.pone.0153417
- 58 Créquit P, Trinquart L, Ravaut P. Live cumulative network meta-analysis: protocol for second-line treatments in advanced non-small-cell lung cancer with wild-type or unknown status for epidermal growth factor receptor. *BMJ Open* 2016;6:e011841. 10.1136/bmjopen-2016-011841
- 59 Brooks-Renney C, Costello S, Kusel J, Timm B. The use of mixed treatment comparisons in NICE technology appraisals. 2011;
- 60 EUnetha. Methodological guideline for REA of pharmaceuticals: Direct and indirect comparison. [cited 2017 Apr 24]. [www.eunetha.eu/outputs/methodological-guideline-rea-pharmaceuticals-direct-and-indirect-comparison](http://www.eunetha.eu/outputs/methodological-guideline-rea-pharmaceuticals-direct-and-indirect-comparison)
- 61 Siemieniuk RA, Agoristas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. *BMJ* 2016;354:i5191. 10.1136/bmj.i5191
- 62 Mann H, Djulbegovic B. Comparator bias: why comparisons must address genuine uncertainties. *J R Soc Med* 2013;106:30-3. 10.1177/0141076812474779
- 63 Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review[review]. *BMJ* 2013;346:f1135. 10.1136/bmj.f1135
- 64 Habre C, Tramèr MR, Pöpping DM, Elia N. Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection. *BMJ* 2014;348:g5219. 10.1136/bmj.g5219
- 65 Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol* 2013;13:50. 10.1186/1471-2288-13-50
- 66 Goudie AC, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *J Clin Epidemiol* 2010;63:983-91. 10.1016/j.jclinepi.2010.01.022

### Supplementary information: additional material