Author’s response to reviews

Title: Network meta-analysis of randomised trials of pharmacological, psychotherapeutic, exercise and collaborative care interventions for depressive symptoms in patients with coronary artery disease: hybrid systematic review of systematic reviews protocol

Authors:

Frank Doyle (fdoyle4@rcsi.ie)
Kenneth Freedland (freedlak@wustl.edu)
Robert Carney (carneyr@wustl.edu)
Peter de Jonge (peter.de.jonge@rug.nl)
Chris Dickens (C.M.Dickens@exeter.ac.uk)
Susanne Pedersen (sspedersen@health.sdu.dk)
Jan Sorensen (jansorensen@rcsi.ie)
Martin Dempster (M.Dempster@qub.ac.uk)

Version: 1 Date: 20 Dec 2018

Author’s response to reviews:

Dear Dr Cheungpasitporn,

We want to thank you for the opportunity to resubmit the following manuscript:

“Network meta-analysis of randomised trials of pharmacological, psychotherapeutic, exercise and collaborative care interventions for depressive symptoms in patients with coronary artery disease: hybrid systematic review of systematic reviews protocol”

We have replied to the 7 reviewers overleaf. We are particularly pleased that 3 reviewers recommended publishing without changes, 2 had minor queries, and we have worked hard to
address the queries of the other reviewers. We have made changes, using the ‘track changes’ option, within the manuscript so that these issues can be easily seen.

Yours sincerely,

Dr Frank Doyle (on behalf of the authors)

Reviewer reports:

Reviewer #1:
The question is interesting. The method of literature review and analysis was valid. Manuscript is well written and clearly explain the process. I have no comments to improve this manuscript.

Thank you.

Reviewer #2:
Thank you for giving me an opportunity to review this manuscript.

This is very informative and well-written manuscript that will establish homogeneous research in this topic of interest.

This is novel manuscript that has been well conducted and constructed. I would happy to support this manuscript to be published and see the outcome of future studies using the protocol from this manuscript.

Worth to be mentioned, many area of knowledge under this topic still needs to be addressed given substantial confounding factor ranging from patients's characteristic to intervention targeting depressive symptoms. I hope this protocol would facilitate more systematic means to conduct the research and then we can implement results from different studies with lessen discrepancy among those studies.

Thank you.
Reviewer #3:

Overall, a competent study plan. However, some important concerns and queries have to be addressed before publication can be advised.

Specific comments:

- Would the scope of this study be too wide and ambitious? "pharmacological, psychotherapeutic, exercise and collaborative care interventions"

We believe that given the number of meta-analyses already published in this area that it is certainly feasible to incorporate them into one review. We have outlined the rationale for this in the protocol. We do not expect to include more than ~30 RCTs in total, so the protocol should not be too ambitious.

- The assessment of the risk of bias and its consideration in the network meta-analysis is significantly more challenging than in conventional meta-analysis. Do you have any strategies in mind to adjust for biases within the included trials?

We agree that the assessment of risk of bias is more challenging for NMA. However, there are several reasons not to conduct moderator analyses based on risk of bias scores. Firstly, we will probably have relatively few studies (i.e. ~30), so moderator analyses will probably be underpowered. Secondly, by their nature psychotherapeutic trials cannot be double-blind, therefore we would automatically be including those as a higher risk of bias – thereby perhaps underestimating their value in any moderator analysis. Instead, we will follow the recommendations for NMA and use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework for obtained results. We have outlined these risk of bias issues within the manuscript in the “Risk of bias and quality ratings” section. We have also proposed to conduct sensitivity analyses with RCTs at low, medium and high risk of bias – outlined in the statistical analysis section.

- Please change "Depression is a common in patients" to "Depression is common in patients".

This has been changed.
- Please rephrase "While a substantial body of research exists in terms of depression interventions in those with CAD [4, 6-11], the effects of both pharmacological and psychological interventions are typically small [6, 8, 12], and may even be smaller than seen in populations without CAD [13], or indeed those with other chronic conditions such as diabetes [14]." The sentence is long, convoluted and confusing.

We have rephrased as follows:

"While a substantial body of research exists in terms of depression interventions in those with CAD [4, 6-11], the effects of both pharmacological and psychological interventions are typically small [6, 8, 12]. Indeed the effect sizes from these studies may even be smaller than those seen in general population samples [13], or other chronic conditions such as diabetes [14]."

- What does it mean by "using the recommended to maximum doses of pharmacotherapies, greater than 4 sessions of psychotherapies - these would be considered high intense interventions; otherwise considered as low-intensity"?

As therapeutic interventions for depression can have different levels of intensity, we intend to generate a binary variable where we can compare high-intensity versus low-intensity interventions. Thus, high-intensity interventions will be those that use the maximum dose of pharmacotherapy, or 5+ psychotherapy sessions. Otherwise, these will be considered as low-moderate intensity. To clarify, we have rephrased this as follows:

“Sensitivity analysis will address different levels of risk of bias (low, medium, high) [58], but also intensity of interventions as rated by a dichotomous variable (as rated by FD and MD; i.e. using the recommended to maximum doses of pharmacotherapies or greater than 4 sessions of psychotherapies – these will be classified as high intensity interventions; otherwise low dosage pharmacotherapy or 4 or fewer sessions of psychotherapy will be classed as low-intensity).”

- It is unclear if the authors intend to report all pair-wise effect estimates together with the associated confidence or credible intervals. This should be clarified, depending on the statistical model used (i.e., frequentist or Bayesian model).
We will report all pair-wise effect estimates, along with confidence intervals, as we are using a frequentist approach. We have added this to the manuscript:

“All pair-wise estimates and associated 95% confidence intervals will be reported.”

- For each treatment, one can calculate the probability that the treatment is the best, second best, or third best among all treatments. However, such probability statements should be interpreted carefully since the difference between treatments might be small and not clinically meaningful.

This is indeed true – we have added this caveat as follows:

“Although we will report the probability that a given treatment is best, second best, third best etc, such probability statements will be interpreted cautiously unless there are actual clinically meaningful differences among the interventions.”

Reviewer #4:

Excellent protocol in an important area. Method choice will give an opportunity to surface evidence in well structured way and the issue of transitivity is dealt with well as are the options for reporting the review using Rankograms which are useful for clinicians (obviously accompanied by the details of effect sizes). No hesitation in recommending for publication

Thank you.

Reviewer #5:

The manuscript SYSR-D-18-00327 presents the protocol for a network meta-analysis that focus mainly on the efficacy and acceptability of pharmacological, psychotherapeutic, exercise and collaborative care interventions for depressive symptoms in patients with coronary artery disease. The protocol describes adequately the population, interventions, comparators, and outcomes for the network meta-analysis as well as the search strategies and statistical analyses plan.
I've only some comments on the manuscript:

- Authors seem to plan to search evidence by filtering results for English language but they also comment on translations for summaries, thus it is not clear if there will be a restricted-by-language search or not. Obviously, if the authors plan a restricted search they should clearly say so and justify it and the likely publication bias (tower of Babel bias) the review could therefore present.

We thank the reviewer for this point. On reflection, we have decided to include only articles published in English (as is typical in such reviews). We will not restrict the search by using language limits, but we will filter out the non-English articles at screening stage. We have changed this throughout the manuscript and added this point about potential bias as follows:

“Searches will not be filtered by language, but non-English language articles will be omitted at the title screening stage.”

“If relevant RCTs are not summarized sufficiently in English in any found review, we will exclude them.”

“Although this may lead to publication bias, exclusion of non-English articles is typical in this field [4,11,12,27].”

- It is not clear if the quantitative outcomes will only be extracted for change scores (end-of-trial minus baseline scores) or if trials with only end-of-trial scores would also be extracted and analysed. Working only with change scores would surely reduce the available evidence to combine.

We intend to use change scores, as stated in the protocol (Section on ‘Study types’): “We will include all randomised trials of interventions for depression in patients with CAD, including pharmacotherapy, psychotherapy, exercise or collaborative care, which use a validated depression scale or diagnostic interviews as an outcome measure, and report a (potential) change in depressive symptoms from baseline or pre-treatment to post-treatment.”

While the reviewer is correct in that this may reduce the available evidence, there are appropriate concerns about combining change scores and end-of-trial only scores (REFERENCE**). We will
therefore consider a supplementary analysis where we include the end-of-trial only scores if this is required to generate the network or there is likely to be substantial missing data (i.e. >10% of trial estimates missing). We have included this as follows in the analysis section:

“As some trials may not report change scores, but may report end-of-trial scores only, we will consider a supplementary analysis where we include the end-of-trial only scores if this is required to generate the network, or there is likely to be substantial missing data (i.e. >10% of trial estimates missing).”

- Authors should also look for clinical trials at clinicaltrials.gov.

We are happy to take this suggestion and have amended the protocol as follows:

“We will also supplement this with a search of clinical trials registries: World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) and clinicaltrials.gov, using similar terms.”

- Authors have described Stata and user-written packages (metaeff, mvmeta, network) as main resources in the analysis plan. However they also quoted R and it would be helpful to know also which libraries the authors will use if needed.

We only intend to use R if required by reviewers at the submission of results stage – i.e. if there is some analyses requested that is not available in Stata. None of the research team are proficient in R, and new NMA packages for R may be written and published prior to the submission of this work. Therefore, we prefer not to indicate the R packages that could be used. Instead, we have deleted this reference to R.

Reviewer #6:
This is very well written protocol and originally written.
Minor comments:

1. "possible different among groups" is not correct in grammar.

2. "in any setting except cardiac rehabilitation' is not correct in grammar.

Thank you – we have changed these as follows:

“Heterogeneity variance will be considered equal within groupings, but possibly different among groups.”

“Participants: patients with CAD and elevated depressive symptoms (clinical diagnosis of depression, or scoring above threshold on a validated scale) enrolled in randomised trials for depression treatment in any setting, excluding cardiac rehabilitation”

Reviewer #7:

The objectives of this network meta-analysis (NMA) are to "compare the best-established treatment(s) for depression in patients with coronary artery disease (CAD) in terms of efficacy (or effectiveness, in studies using non-placebo comparators) and acceptability, with several secondary outcomes also considered." The proposed primary outcomes are change in depressive symptoms (summarized with standardized mean difference (SMD)) and treatment acceptability (treatment discontinuation: % people who withdrew). Secondary outcomes include change in 6-month depression outcomes, health-related quality of life (HRQoL), mortality, cardiovascular morbidity, health services use, and adverse events. The authors propose this review as the first NMA to provide a ranking of treatments for depression in those with CAD. However, there are a number of methodological concerns about the conduct and reporting of a network meta-analysis that should be addressed. In addition, the review protocol is not reported in many places with sufficient detail or consistent with PRISMA.

Major Comments

1) The rationale for excluding cardiac rehabilitation because it is "currently undergoing investigation elsewhere" is inconsistent with the logic of conducting a NMA for depression treatments. The cited cardiac rehabilitation meta-analysis (reference 33) is not about treatments for depression. If cardiac rehabilitation is considered a depression treatment alternative to those
included in the present NMA, then cardiac rehabilitation should be included. If it is not an intervention that would be used for treating depression, then it should be excluded. The authors should clarify this rationale.

We appreciate the reviewer’s concern here. While the cited NMA for cardiac rehabilitation does have depression as a secondary outcome of the review, the reviewer is correct that cardiac rehabilitation, per se, would not be considered a front-line treatment for depression – and this is a better justification for omitting it. We have therefore re-worded the manuscript as follows:

“Cardiac rehabilitation will not be included, as this is currently undergoing investigation elsewhere [33], but more importantly because cardiac rehabilitation per se would not be considered a front-line depression intervention.”

2) The authors state that they will exclude RCTs of “psychological interventions that are not established psychotherapies delivered by trained therapists.” How will this be codified for inclusion/exclusion?

As per the manuscript, in the inclusion criteria we state we will include “Psychotherapy trials delivered by trained therapists: cognitive-behavioural therapy, interpersonal psychotherapy, mindfulness, etc.” We will therefore codify this for inclusion as follows:

- Talking therapies (yes/no)
- Trained therapists (yes/no)
- One of the following psychotherapies (Yes/No; as per Dragioti et al (2017) “Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials” Acta Psychiatr Scand, 136, 3, 236-246.):
  - Acceptance and commitment therapy
  - Behavioral therapy
  - Cognitive Behavioral Therapy
  - Cognitive therapy
  - Cognitive remediation
Cognitive stimulation therapy
- Dialectical behavior therapy
- Family systems therapy
- Integrative psychotherapy
- Interpersonal psychotherapy
- Mindfulness-based therapies
- Multimodal therapy
- Positive psychology interventions
- Problem solving therapy
- Psychodynamic psychotherapy
- Supportive therapy and counseling

We have added this list to the paper, as follows:

“Psychotherapy trials delivered by trained therapists: cognitive therapy, cognitive-behavioural therapy, interpersonal psychotherapy, mindfulness, acceptance and commitment therapy, behavioral therapy, cognitive remediation, cognitive stimulation therapy, dialectical behavior therapy, family systems therapy, integrative psychotherapy, multimodal therapy, positive psychology interventions, problem solving therapy, psychodynamic psychotherapy, supportive therapy and counseling.”

3) Will the authors include abstracts? Will they seek data from unpublished studies (they say they will search trial registries)?

We included the following in the original submission (pg 13):

“Unpublished data will be requested for unpublished or ongoing studies.”

However, as randomised trials usually have locked-down datasets prior to primary outcome analysis, we are not hopeful of obtaining much data in this way. We therefore will not include data from abstracts only:
“Unpublished data will be requested for unpublished or ongoing studies, but data from abstracts only will be excluded.”

4) The authors state that they will do a review of systematic reviews to identify eligible RCTs for their NMA. Per the Cochrane Handbook (https://handbook-5-1.cochrane.org/chapter_22/22_overviews_of_reviews.htm), however, an overview of reviews is a synthesis of systematic review results. It would be reasonable for the authors to use existing systematic reviews as a starting point for a systematic review and NMA of primary RCTs. However, that would require a methodology for determining whether identified systematic reviews were done adequately (e.g., per AMSTAR) and would require a plan for updating to determine if there are more recent relevant trials. This would be important given the relatively small number of trials, for instance in drug interventions, and the possibility that a newer trial could feasibly change effect estimates.

We are glad that the reviewer states it is reasonable for to use existing systematic reviews as a starting point for our NMA. The reviewer is correct that, strictly speaking, an overview of reviews methodology concerns inclusion of reviews only. However, NMA is far more powerful when the original trials are included, than when simply the summary estimates of meta-analyses are included – indeed including original trials is highly recommended. As we propose to obtain original RCT summary estimates, we therefore do not need to use AMSTAR, but are instead using risk of bias assessments for the original RCTs. We also do include a plan for updating the reviews – we clearly state this in the original manuscript – that we will search the past 5 years for newer RCTs.

However, we appreciate the reviewer’s methodological concerns, so we have modified the language in the manuscript to indicate that we are adopting a hybrid approach, as follows:

Title: “Network meta-analysis of randomised trials of pharmacological, psychotherapeutic, exercise and collaborative care interventions for depressive symptoms in patients with coronary artery disease: hybrid systematic review of systematic reviews protocol”

Search strategy and study selection: “As several systematic reviews have already been published in this area (e.g. [6-11]), we will conduct a hybrid overview of reviews [31] and systematic review methodology. We will find and extract RCTs and their associated data using the content of these reviews and the original RCT papers (including RCTs which are listed as not having being included in the final published reviews (e.g. [6, 9]).”
5) Setting clear PICO elements is very important in NMA to ensure that the key assumption of transitivity is likely to hold. In this review protocol, the authors may need to narrow down the criteria for the eligible participants and included interventions and clarify under what conditions the transitivity assumption is likely to hold.

We thank the reviewer for this point. We have narrowed the inclusion criteria for psychotherapy RCTs, as per the point above. We do not wish to modify the CAD inclusion criteria as treatments for CAD are similar across acute coronary syndrome, angina, angiographically confirmed coronary disease, receipt of percutaneous coronary intervention or coronary bypass graft. We do not wish to modify the depression inclusion criteria as these are typical for the currently published systematic reviews in this area. We have described in detail how the transitivity assumption will be addressed within the manuscript. We have added the following, to provide the clarity the reviewer requests:

“If transitivity is not demonstrated (e.g. if there are clear, statistically significant and/or clinically important differences in patients enrolled to trials in terms of age, sex, CAD or depression severity indices [48]), we may explore building separate networks to reflect the evidence.”

6) In the PICO statement that the authors present, participants of the review are patients with CAD and elevated depressive symptoms (clinical diagnosis of depression, or scoring above threshold on a validated scale). However, they do not specify the clinical diagnosis of depression, validated scale, and the threshold. Does any clinical diagnosis that is made by a clinician count as a valid diagnosis? Which scales are validated? Do the studies using the same validated scale apply the same threshold? Examples should be provided for clarification.

Yes, we will include any diagnosis of depression made by a clinician. There are potentially dozens of validated depression scales that could be used, but typically the following are used: Beck Depression Inventory, Hospital Anxiety and Depression Scale, Center for Epidemiological Studies Scale for Depression, Symptom Checklist 90-Revised, Patient Health Questionnaire-9, Hamilton Depression Rating Scale etc – however, it does not seem useful to cite all of these at this stage. It is unlikely that all studies will use the same thresholds, but this important aspect will be dealt with as per the transitivity assumption as above. We are happy to clarify the general points as follows within the manuscript:
“Participants: patients with CAD and elevated depressive symptoms (clinical diagnosis of depression [any clinical diagnosis of depression that is made by a clinician or by structured diagnostic interview], or scoring above threshold on any validated depression scale) enrolled in randomised trials for depression treatment in any setting, excluding cardiac rehabilitation”

7) Another part of the PICO statement is the comparison. In this review protocol, comparison groups include placebo groups, usual care or waitlist control, attention control groups. Although the authors provide a rationale for the inclusion of several comparison groups, the proposed network plot is confusing. That is, since there are multiple comparator groups in the entire network, how will a single treatment ranking be generated?

The treatment ranking will simply include all the different groupings outlined in the protocol. For example, it is possible that the treatment rankings will be as follows (worst to best): usual care, placebo, minimal treatment control, antidepressants, exercise, psychotherapy, collaborative care, combination therapies.

8) Regarding the primary and secondary outcomes, the authors do not provide the rationale for the time windows they choose. For instance, why the week 8 post-intervention (between 4-16 weeks)/week 26 (between 20-30 weeks) is selected as the time point for SMD estimates as the first primary/secondary outcome of efficacy(effectiveness)-response (continuous)? In addition, why the range of 4-16 weeks is selected for the last secondary outcome, adverse events?

We did provide the rationale, as follows:

“For the synthesis of primary outcomes and the secondary outcomes of adverse events, we will adopt the 8-week threshold as per previous review, or, if data is unavailable for this duration, we will use the closest available data from 4-16 weeks) [13, 36]. This was adopted as researchers believed that depression treatments should work in a clinically reasonable period of time.”
9) The search string is appended in this article, but it is adapted from an earlier work of Baumeister et al. The authors are encouraged to have the search strategy peer reviewed via the PRESS system.

We thank the reviewer for this suggestion – we have consulted a n independent librarian who has agreed to review the search using the PRESS checklist.

10) In Data Extraction for Continuous outcomes, the authors mention that missing imputation with the median from the other studies within a same group will be used if there is insufficient data for calculating 95% confidence interval. They state that a sensitivity analysis will be followed. They need to provide rationale for their decision. Multiple imputation, accounting for the uncertainty of the imputed value, may be a better approach.

The reviewer is correct that multiple imputation techniques are the best way of managing missing data. The difficulty is that it is currently unclear whether the multiple imputation techniques available within Stata (the statistical package we are adopting, using the Stata-developed – mi- commands) will be implementable with the user-written network analysis (which involves the user-written –mvmeta- and -network- commands). Also, we think that it is unlikely that there will be >5% missing data for calculating the 95% CI – and in these instances there are limited benefits for using the more sophisticated missing data techniques. We have addressed the reviewer’s concern as follows:

“If insufficient data is available to calculate the 95% confidence intervals, we will impute this will the median from the other studies within that particular grouping. If >5% of studies require such imputation, we will consider multiple imputation techniques.”

11) In Data Extraction for Missing RCT outcome data, the authors claim that "the missing data imputation is already part of the risk of bias assessment conducted in prior meta-analyses", however, the review they propose here actually includes both studies involved in prior meta-analyses and "RCTs which are listed as not having being included in the final published reviews". The authors need to specify how they would address the missing RCT outcome data in the newly-added studies.
We stated in the protocol that (Cochrane) risk of bias assessments will be conducted by the authors where these were unavailable in prior systematic reviews – i.e. in the more recent RCTs that will be found with the supplementary search strategies (see above).

12) In the statistical analysis section, it is unclear that how the multiple comparator groups issue is addressed in the main network analysis. Would comparison between different comparator groups, between placebo group and minimal treatment group, for instance, be included in the main network?

Yes, we are including these comparator groups separately in the main network, as illustrated in Fig 1. However, it is unlikely that many trials will have more than one comparator group (e.g. placebo and usual care). We have clarified this as follows:

“Several comparator groups will be used in the main network, as outlined in the ‘Comparison groups’ section above.”

13) Do the authors have an a priori idea of how many eligible studies they will find for the different included interventions for CAD (excluding other heart disease patients)? That is, have they done preliminary work to ascertain the likelihood that they will be able to accrue enough RCT data to draw conclusions with any confidence? What can be understood from existing systematic reviews on included interventions.

As mentioned above, we do not expect more than ~30 RCTs. Whether this network of evidence is enough to draw conclusions with sufficient confidence is open to debate, and will be dependent on a the final results, which could be influenced by several factors; including, but not limited to: number of studies in each grouping, the variability of the effect sizes, the individual precision of these effect sizes, etc. Even if we do not have much confidence in the ultimate results, we can make this clear in the final submitted manuscript – and this will then become an important issue for future work. As per the reply to other reviewers above, we have already included information from prior reviews – such that the effect size for depression interventions post-CAD is smaller than would be expected, etc.
14) For a key assumption of NMA, the authors make a plan to assess transitivity and list potential effect modifiers. "differences in placebo-controlled versus other comparator group studies" is one of the effect modifiers, but it is unclear that how these will be measured and how the differences will provide evidence for transitivity. In addition, does this mean only placebo-controlled groups are used as the comparison group in the main network? The authors need to clarify.

The reviewer may have misunderstood this. As outlined above and in the manuscript and fig 1, we are classifying different comparator groups into 1) pill placebo; 2) no treatment, waitlist or treatment as usual; 3) minimal treatment control, active comparator, specific and non-specific factors treatment control. These will all be included in the main network. In NMA, all groupings are compared to each other, including all the treatments and the comparator groups – we are not necessarily limited to once comparator group such as placebo.

15) There are some areas where the reporting is not consistent with PRISMA-P and PRISMA extension statement for NMA requirements that need improvement (this is a protocol, items for reporting results and discussion are not applicable). Some examples are below.

a. Item #2: the authors are encouraged to register the protocol first because there are a large body of systematic reviews and meta-analysis in this area, there might be similar NMA already be registered.

We thank the reviewer for this – we have already registered the protocol and have provided the registration number as follows:

Abstract: “Systematic review registration: CRD42018108293 (PROSPERO)"
Method: “This protocol has been registered on PROSPERO (CRD42018108293)”

b. Item #16: Meta-bias, for example selective reported within studies is mentioned on page 16, but no specific plan of assessment is provided.
This was included in the original submission (page 19 of current submission):

“Funnel plots for NMA, which plot the difference between the study-specific effect sizes from the corresponding comparison-specific summary versus the inverted standard error, will be used to ascertain whether estimates from more imprecise RCTs are different from those RCTs with more precision (such as larger effect sizes for depression treatment in smaller studies) [28, 50]. A network meta-regression will investigate associations between effect size and study sample size.”

16) The page numbers for reporting many of the items in the PRISMA-P and PRISMA extension statement for NMA requirements (the page numbers are different even referring back to the original word file they submitted) are not consistent with the main document text. The authors should double-check and ensure the readers can locate the page number correctly. For example:

a. Item #3b: contributions statement is on page 21, not 12.

b. Item #5a: Financial statement is on page 21 too, not 12.

c. Item #9: Information sources is on page 12-13, not 7-8.

d. Item #12: Date items are on page 12-13, not 8-9.

e. Item #14: Risk of bias assessment is on page 15-16, not 8-9.

f. PRISMA NMA Item #S1: Geometry of the network is reported on page 16-17 not 10-11.

g. PRISMA NMA Item #13: Summary measures is mentioned on page 19, not 11.

We thank the reviewer for this, which occurred due to the submission of an incorrect checklist from an earlier version of the manuscript. These have all been corrected on the checklist. We apologise for this error.
17) In the statistical analysis section, the authors describe the commands they will use to perform the NMA. In addition to this information, the authors should describe the statistical models that underlie these commands, and their assumptions.

While describing the full statistical models and assumptions are not usually part of systematic review protocols, we have provided summary detail with reference to more detailed papers for the interested reader, as follows:

“Two main networks will be evaluated using frequentist multivariate meta-analysis (commands network meta and mvmeta, which underpins the first command) in Stata 15 [51]. These commands perform restricted maximum likelihood methods for random effects multivariate meta-analysis by using a Newton-Raphson procedure, accounting for within- and between-study correlations. The assumptions of this model are that the multiple modelled effects represent a multivariate normal approximation of the estimated effects; that a multivariate linear regression can be performed due to linear effects between studies; a constant between-studies covariance matrix, where conditional variances of all components of the random effect are constant; a symmetrical normal distribution which does not allow for light or heavy tails (which consequently means that outlier trial results can be overly influential for final estimates) [51]. The interested reader is referred to the following references for more detail [51, 52].”

18) In the statistical analysis section, the authors describe pairwise meta-analyses. Please provide a clearer rationale for these analyses.

We have provided this as follows:

“Pair-wise meta-analytic estimates are usually reported in addition to the network estimates [13, 36, 52]. Among other reasons, they are useful to determine 1) the potential effects of any outliers, and 2) to demonstrate any differences in estimated effects from the network meta-analysis which could be attributed to correlation between the outcomes – which is ignored in pair-wise meta-analysis [52].”
Minor Comments

19) The authors do not check any item in the column of "Information reported" in the checklist of PRISMA-P and PRISMA extension statement for NMA requirements.

We have now addressed this as per the comment above, by including the correct checklist.

Changes not requested by reviewers:

We made 3 changes to the protocol that were not requested by reviewers:

1) As final searches will now be conducted in 2019 (not 2018), we have edited the search strategy to reflect this:

“Then we will conduct a search of MEDLINE/PubMed and the Cochrane Library for recent RCTs published in the last 5 years (i.e. from 1st January 2014).”

2) We will no longer be in a position to obtain translations of different articles, so we plan to include English-language articles only:

“If relevant RCTs are not summarized sufficiently in English in any found review, we will exclude them.”

3) We have modified the inclusions of adverse events (and/or dropouts), due to concerns that these may not usually be reported within the 4-16 week timeframe:
“If data on adverse events (dropout) is not available for this time period, we will consider using the overall dropout rate as a proxy.”