Author's response to reviews

Title: Comparative Effectiveness of Psychological Treatments for Depressive Disorders in Primary Care: Network Meta-Analysis

Authors:

Klaus Linde (Klaus.Linde@tum.de)
Gerta Rücker (ruecker@imbi.uni-freiburg.de)
Kirsten Sigterman (kirsten@sigterman.nl)
Susanne Jamil (susanneohnsmann@gmx.de)
Karin Meissner (Karin.Meissner@med.uni-muenchen.de)
Antonius Schneider (antoniusschneider@tum.de)
Levente Kriston (l.kriston@uke.uni-hamburg.de)

Version: 3
Date: 8 July 2015

Author's response to reviews: see over
Responses to reviewers *(in italics)*

We thank the reviewers for their careful and constructive comments!

General response to reviewers 1 and 2

It seems that our original manuscript did not make sufficiently clear how much larger the database is in the network meta-analysis presented compared to our previously published conventional meta-analysis of usual-care controlled trials. In fact, in the current analyses data from 37 studies with 7,024 patients were included, which is approximately a third more than the data analyzed in the previous conventional meta-analysis (30 studies with 5,159 patients). We tried to correct this by introducing the differences shortly in the introduction, by reporting the number of comparisons used in the results, and by addressing these differences more explicitly in the discussion – see also our responses to your comments.

Reviewer 1 Frank Doyle

The authors conducted a network meta-analysis of trials comparing psychological treatments for depression. The study is well-conducted, but the authors need better justification regarding some specific points. The only major issues for me are whether the amount of detail on the standard meta-analyses is required/ justified – since this appears to have been published already? I am also unclear as to how the OR and SMD data was combined, if it was? Or indeed the secondary outcomes? Details below.

*R – justification for presenting detailed standard meta-analyses/secondary outcomes:* Network meta-analyses integrate a huge amount of data into a single forest plot (e.g. Figure 2) or table (Tables 2 and 3). In such summary displays, only pooled estimates are presented without any information on single studies. We think that the crude data used to calculate effect size estimates of the single studies included into (pairwise and network) meta-analyses and the resulting effect sizes should be transparent and available to readers and other researchers, like recommended by relevant reporting guidelines (Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Int Med 2009; 151: 264-269; Hutton B et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015; 162: 777-784). Our previous publication in the Ann Fam Med 2015; 13:56-68 contains only single study data of comparisons with usual care with for a single outcome (post-treatment depression scores). So, at best one might discuss whether the forest plot 4.3 in the appendix is necessary – but even this plot contains data not included in our previous publication (findings for SSRI and TCA). All other 19 forest plots in the appendix contain single study data not included in the previous publication; these data are the basis for the network meta-analysis on all four outcomes. The network meta-analysis findings for ALL outcomes are actually presented in Tables 2 and 3. As in a number of other network meta-analyses, these tables present different outcomes above and below the diagonal. We now explicitly point to this when referring to tables. Please see also our general response above.

*R – combination of OR and SMD:* we did not combine OR and SMD data. We agree this would have been possible (and maybe wise). Yet we had not planned it in our analysis plan. We also think this
should be acceptable as we had data on each response, remission and discontinuation for 34 of 37 studies (partly by imputing dichotomous outcomes from means and standard deviations), and for 36 of 37 studies for post-treatment scores (see Table 1).

Discretionary revisions

Abstract

The Background sentence is an aim, not the background?

R – A new first phrase has been added (background is the standard heading in BMC Fam Pract abstracts).

Pharmacological treatments are reported in the Methods, but not in the abstract?

R – We did not present results on pharmacotherapy a) to keep the abstract concise and b) as the important comparison with drugs is difficult to interpret due to the small number of trials.

Minor essential revisions

Background

It is unclear from the authors description whether the Barth meta-analysis (ref 5) included or excluded studies which had reduced HCP time with patients?

R – The Barth meta-analysis included such trials but it did not address the problem in a manner comparable to our review. The only analysis performed in respect to kind of treatment format and delivery was to compare individual face to face therapies vs. all other (differences were not significant). This separation means that face to face group therapies were analyzed together with remote/limited/no contact therapies. The number per sub-category among “other” is unclear. Together with the preponderance of studies performed in specialized care (where very low-contact studies seem unlikely), this makes it very difficult to compare these analyses with our findings. We would prefer not commenting on this complex issue in the introduction, but have added it in the discussion where we compare our findings with those of other research (lines 410 to 414 in the track change version).

Line 105 – typo ‘of’ should be or?

R – corrected

Methods

The last update searches occurred in December 2013 – could or should these be updated again, given that this is now 2015?

R – We agree that it would be desirable to update the search. However, this would imply that all analyses would have to be redone. Unfortunately, we do not have the resources to do this. Funding of
this study ended in spring 2013 (but due to the complexity completion of the review took much longer).

Results

Why isn’t the flow chart part of the results rather than a supplemental piece? The flow chart does not indicate how many articles were chosen as a result of reviewing other authors meta-analyses (i.e. how many, if any, articles were missed by the authors search terms and execution)?

R – We presented the flow chart in the appendix (which is not unusual in network meta-analyses) to limit the number of figures and tables, particularly, as tables 2 and 3 are very large. If desired by the editor, the flow chart can easily be moved into the main manuscript. Frankly, we cannot easily say in a definitive manner how many of the finally included studies were identified by reference tracking, as potentially eligible articles were added to the reference database earlier in the process (and together with drug trials for a parallel review). Yet, we can clearly say that no more than three of the included trials remained unidentified by the database searching.

The HADS would be an unusual outcome measure in such trials, and our work shows that it is of questionable validity (e.g. Cosco, T. D., et al. (2012). "Latent structure of the Hospital Anxiety And Depression Scale: A 10-year systematic review." Journal of Psychosomatic Research 72(3): 180-184.) Was there no other measure for this study (Naeem 2010)?

R – The only other outcome measure used was the Bradford Somatic Inventory which we considered less appropriate. The HADS might have been used as it was available in the language needed (Urdu). Furthermore, the HADS was shown to measure severity of depression comparably to several other depression questionnaires (Wahl I et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. J Clin Epidemiol 2014; 67: 73–86).

There is a typo in supplemental 1 Table 3.2 - HAD-D = Hospital Anxiety and Depression Scale subscale depression.

R – corrected (thank you!)

Major revisions

Methods

Lines 169-70 – I think it would be better to supply an example for readers of how responder/remission data was imputed.

R – We modified the phrase to make explicit what data was used for imputation: “In cases where responder and remission data was not reported it was imputed from means and standard deviation using the method described by Furukawa et al. [23]. For response imputation, first the threshold of at least 50% score reduction after completion of treatment was defined by halving the mean baseline score. Then, the proportion of participants reaching this threshold was calculated assuming a normal distribution of the data after completion of treatment. Further details, examples, and empirical justification are given by Furukawa et al. [23] and Meister et al. [24]”.
Analyses

It would be useful to know why the authors did not transform ORs to SMDs, in order to combine all the data? This approach has been recommended and implemented before

*R – please see our answer to the very first comment*

While these are discussed separately (ORs and SMDs), I do not see separate results for each method if they were treated separately?

*R – Table 2 and 3 provide for all 4 outcomes the findings for all possible comparisons (please see also our answer to the very first comment)*

It would be useful to describe in more detail ‘consistency models’ etc – not all readers will be familiar with these terms? I am not.

*R – We have revised and amended the text substantially in order to enhance comprehensibility.*

Overall, I think this section could do with more detail as there are no space constraints for this journal?

*R – We hope that our additions and modifications made the section more understandable.*

It is unclear why the Forest plots for direct comparisons are being reported, if they have been reported previously? The order of the comparisons changes – favours treatment/controls changes from 4.2-4.3 – why? If these are to remain, along with the funnel plots, they should be referred to in the Results.

*R – We apologize for the switched order in forest plot 4.3 (corrected). Regarding forest plots, we revised the text in order to make it more clear that all forest plots are based on network meta-analysis estimates, not direct estimates. Please see our answer to the very first comment. We refer now to the funnel plots in the text.*

232 - How exactly did citation 29 disagree with the main body of evidence?

*R – The source of this result is the inconsistency analysis based on changes of the between-designs Q statistic after detaching of single designs (ref [25] of the manuscript), illustrated by the net heat plot (supplement, Section 7). Inconsistency can be located in the comparison PST vs UC of study [29] with a direct estimate of an odds ratio of 4.01 [1.37, 11.80] compared to the network estimate which is 1.40 [0.99; 1.98]. That is, the study showed an exaggerated effect of PST vs UC compared to direct and indirect evidence from other studies in the network. We added this explanation to the manuscript on page 10.*
I may be missing something, but the authors wished to examine the effects for the outcomes response to treatment (primary) and remission and post-tx scores (secondary). Where is the NMA plot for the secondary outcomes?

*R – please see our answer to the very first comment for the first point.*

**Reviewer 2 Jane Gunn**

The methods appear appropriate and they are well described. For the non-expert in the area of network analysis an explanation of the terms ‘frequentist method’, consistency model, inconsistency model would be helpful.

*R – We have revised and amended the text substantially in order to enhance comprehensibility for the non-expert readers.*

The authors acknowledge that this work builds upon a recently published meta-analysis in the Annals of Family Medicine (ref#11)– but this could be stated even more clearly and earlier in the paper and the authors should point out how this analysis differs from that already published work and why it is justified.

*R – The whole second part of the background section tries to explain why we believe that the methods and findings in the manuscript go clearly beyond the published standard meta-analysis comparing psychological treatment with usual care. We listed several reasons in the background section of the revised manuscript to emphasize the necessity of such an analysis (including making use of substantially more data, formal estimation of comparative effects, increasing precision of estimates, and being able to check the consistency of evidence). Please see also our general response to reviewers 1 and 2 above.*

The information about the three assumptions that need to be met for a network analysis is very helpful. I think this information would fit better being introduced in the methods section and clearly stating at that point the concerns that the authors had around the transitivity assumption and stating clearly their argument for why they continued with the analysis despite their concern.

*R – This comment addresses a difficult issue which is often neglected in publications of network meta-analyses. We have moved the explanation of the three assumptions into the methods section and directly link it now with the respective assessment methods. We tried moving the findings and the discussion of the assessments into the methods section, too, but found this counter-productive. Addressing the issue completely in the methods section is problematic as it heavily relies on actual results of the network meta-analysis and judgment. Findings regarding homogeneity/consistency are presented in the results section and the appendix. Assessment of transitivity is based on clinical and epidemiological judgment (which is now clearly stated in the methods section). We felt that integrating the summary of findings and interpretation of the three assumptions in the discussion is the clearest way to address the issue (even if the proposal of the reviewer is fully understandable). We hope our current compromise is acceptable.*
After line 323 I think it would be helpful to include a paragraph which discusses the findings in this paper in the light of the findings from their conventional metaanalysis published recently in Annals of Family Medicine (ref #11). From my reading of both papers – the findings are complementary – so what does this new analysis add? The Tables in ref #11 are very similar to those in the Suppl. files – it will be important that the difference between these two papers is clearly explained.

R – We try to make the important differences in methods and data used more clear now (please see also our responses above). The only table (3.2.) in the on-line material similar to our previous publication contains the seven additional trials. We believe that the table must be included – otherwise interested reviewers would find here only the seven additional trials and would have to go to the other paper to find information of the other 30 trials. Please see Item 20 and Appendix Table in the PRISMA Extension statement for Network meta-analysis (Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thorlund K, Jansen JP, Mulrow C, Catalá-López F, Gøtzsche PC, Dickersin K, Boutron I, Altman DG, Moher D.: The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. Ann Intern Med. 2015 Jun 2;162(11):777-84. doi: 10.7326/M14-2385).

The authors note that for some of the comparisons they had little data. Can they make a comment about the IPT data- as they state this intervention has conflicting evidence – do they think that they have had sufficient data to generate a reliable finding?

R – We have added corresponding information to the discussion section concluding that our confidence in the findings on IPT should be considered moderate.

The finding of reduced effectiveness for the minor severity disorders is interesting. Do the authors consider “reducing depression severity” as a good outcome measure for judging effectiveness in this group of disorders? Do they think that alternative outcome measures might be more appropriate? Depending upon their view, they may wish to comment on this in the discussion.

R – This is an interesting question. However, we think we cannot answer it based on the analyses we have done. Therefore, we would prefer to not address it.

Line 342 – I suggest they delete the comment regarding CBT and SSRIs’ – as they state they did not have enough data to draw a reliable conclusion and their comment as it stands could lead the reader to draw a conclusion that is not supported by the data presented.

R – We revised the text moderately in order to make more clear that the findings on this treatment are not reliable. We write „Our analyses suggest a marked superiority face-to-face CBT combined with SSRI treatment over all other treatments, but this extraordinary finding from only one small trial should be interpreted with greatest caution unless confirmed by large trials. Clearly, the available evidence does not allow any firm conclusions on the value of combination treatments in depressed primary care patients.”
Line 353 onwards. This section raises the issue of whether usual care was consistent across the studies. What can they say about this? What limitations does it bring?

\textit{R – It clearly was not and we have addressed this in lines 384 to 386 (track change version). In the paragraph between lines 439 and 455 (track change version) we suggest that future studies should not primarily focus on comparisons with usual care.}

It seems to me that the major finding of this review is captured at line 332-333 – that CBT is effective and that it appears to be just as effective when delivered remotely and using minimal contact with health professionals. This finding could be given more prominence and appear in the Abstract? The authors could discuss this finding in more depth: how strong do they consider their evidence is for this finding. Is it strong enough to support health care reform or do we require further studies? Discussing the health care policy implications of this finding would enhance the paper.

\textit{R – We have emphasized this finding more strongly in the conclusion and the abstract of the revised manuscript. Furthermore, we amended the discussion with a judgment on the quality of evidence and further implications.}

\textbf{Reviewer 3 Areti Angeliki Veroniki}

\textbf{Minor Compulsory Revisions}

\textbf{Statistical analyses}

1. Lines 173-177. The authors apply pairwise meta-analyses in a frequentist framework using RevMan, a NMA for binary data in a Bayesian framework using R and R2WinBUGS, and a NMA for continuous data in a frequentist framework using R and netmeta. Why did the authors apply their analyses in different settings? I would suggest applying all analyses in the same framework (frequentist or Bayesian) and using the same estimators for heterogeneity, so that the NMA and pairwise MA results can fairly be compared.

\textit{R – We agree that this might be confusing and is a point for discussion. We give two arguments for explanation, a historical one and a methodological argument. First, when we started this project, the only available method well described and frequently used in the literature was the Bayesian approach using WinBUGS. Particularly, no R package was available at that time. For this reason, in our protocol (ref [14] of the paper) we planned to use the Bayesian approach for our analyses and used the NICE series of documents (now published in Medical Decision Making) that helped us a lot. While we analysed our data, one author (GR) developed a frequentist approach to network meta-analysis, and at the same time others published related methods, e.g., Krahn et al., ref. [25] of the manuscript. We started a collaboration with these authors and collected the frequentist methods in a new R package netmeta, in July 2013 released on CRAN (http://cran.at.r-project.org).

From the methodological point of view, the methods implemented in RevMan and in R package meta are just identical. The method implemented in netmeta, weighted least squares estimation, is a simple generalization of pairwise meta-analysis (as implemented in meta) and again gives identical results for the special case of pairwise meta-analysis. Thus, the only remaining potential difference in modelling is between netmeta and WinBUGS. The reason why our last analysis, that of SMDs, was done with netmeta was that we had always experienced very similar results when comparing both approaches. As we find our R package quite easy to use, user-friendly (e.g., there is no need to change treatment labels or original study labels to consecutive numbers) and much faster than the MCMC
algorithm, we did not see any reason to repeat our analysis using WinBUGS. Moreover, netmeta offers heterogeneity and inconsistency diagnostics based on the decomposition of Q, visualized by the net heat plot [25]. On the other hand, admittedly, netmeta (at the moment) does not allow including covariates in the model. Therefore, for subgroup analyses and meta-regression we continue to use WinBUGS. We do not think that methodological diversity necessarily is a problem. Senn et al. also concludes that the “choice of implementation of framework for inference, Bayesian or frequentist, may be less important than other choices regarding the modeling of effects” (Senn et al. Issues in performing a network meta-analysis. Stat Methods Med Res 2013; 22: 169-189; p. 183).

2. Please provide the priors or estimators used to estimate heterogeneity. Several methods exist to estimate the heterogeneity variance with different properties, and it has been shown that different estimation methods may considerably impact on the meta-analysis results (see for example Cornell et al, Annals of Internal Medicine 2014, Kontopantelis et al PLoS One 2013, Novianti et al Contemporary Clinical Trials 2014).

R – For the Bayesian analysis, we used the code provided in Example 1(c) of Dias 2013 (ref [24] of the manuscript) with a vague uniform (0,5) prior for the between-study standard deviation. For the frequentist analysis, the generalized method-of-moments estimate was used as implemented in netmeta (Jackson D, White IR, Riley RD (2012) Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Statistics in Medicine 31(29):3805-3820, eq. (B5)). We added this information to the manuscript (page 7).

3. Line 182. Citation is needed for the applied inconsistency model. It is not clear to me which inconsistency model was applied (e.g., the model suggested by Lu and Ades J Am Stat Assoc 2006, White et al Res Synth Methods 2012, or the model without consistency constraints (applying pairwise meta-analysis for each comparison, which share a common heterogeneity parameter)).

R – The inconsistency model was the unrelated mean effects model used in ref [27] of the manuscript where WinBUGS code is provided in the appendix. We added this information to the manuscript (page 8).

4. Line 187. Did the authors assess for funnel plot asymmetry using a statistical test apart from the visual inspection of funnel plots? Also, the Cochrane Handbook recommends evaluating funnel plot asymmetry “when there are at least 10 studies included in the meta-analysis” due to the insufficient power of the tests to distinguish chance from real asymmetry.

R – We added the phrase: “As there were no direct comparisons with ten or more trials we did not perform statistical tests to distinguish chance from real asymmetry (insufficient power of such tests)”.

5. Line 231. Is this I2 referred to the heterogeneity for pairwise meta-analysis or to the heterogeneity/inconsistency for network meta-analysis (as suggested by Jackson et al Stat Med 2014)? Please provide a reference and clarify if this is for inconsistency, heterogeneity or both.

R – The I2 referred to that produced by netmeta. Like the univariate I2 (Higgins and Thompson 2002) It is derived directly from Q and measures total heterogeneity and inconsistency (I2 = 100%*max(0,(Q – df)/Q, see eq. (6) in Rücker G and Schwarzer G (2014) Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. Stat. Med. 33:4353-4369. We added this information and the reference (page 7).
6. Line 237. Please change CIs to Credible Intervals, if these have been estimated in a Bayesian framework.

R – We have checked the manuscript completely in order to make sure that we clearly distinguish between the two concepts.

7. Lines 263 to 265. The visual inspection of funnel plots suggested that small trials tend to report more positive findings in some cases. Did the authors also attempt to apply a fixed-effect model and compare the results with the random-effects model? In the presence of small-study effects, the fixed-effect model would probably be best to apply, as small studies are given a smaller weight and contribute less to the overall pooled effect.

R – Yes, we also estimated fixed effect models for all outcomes. As heterogeneity/inconsistency was low for all outcomes, fixed effect and random effects models provided very similar results throughout. We added this remark to the manuscript (page 11). A prevailing problem with the funnel plots here is that they consist of all treatment-placebo contrasts. There is only poor information per contrast, therefore they are of limited informative value anyway.