

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The road to resilience: a systematic review and meta-analysis of resilience training programs and interventions.
AUTHORS	Joyce, Sadhbh; Shand, Fiona; Tighe, Joseph; Laurent, Steven; Bryant, Richard; HARVEY, SAMUEL

VERSION 1 – REVIEW

REVIEWER	Matthew Gallagher University of Houston, USA
REVIEW RETURNED	15-Jun-2017

GENERAL COMMENTS	<p>Review of bmjopen-2017-017858: The road to resilience: a systematic review and meta-analysis of resilience training programs and interventions. This manuscript meta-analytically reviews literature on RCTs that involved resilience programs and certain resilience outcomes. As requested, my review primarily focuses on the methodological features of the study. The authors appropriately follow the PRISMA guidelines in reporting the findings of their meta-analysis and the methods of the literature search and statistical analyses generally appear to be appropriate and conducted properly. There were some areas, however, where it would be beneficial to be more explicit about various methodological details of the meta-analysis. Specifically:</p> <p>*It would be helpful to be more explicit in the methods section on page about was considered a validated measure of resilience and therefore met eligibility criteria for this meta-analysis. As the authors note, the definition and measurement of resilience is very complicated, so being very clear about what was considered an acceptable resilience outcome is crucial for understanding the scope and focus of this review. The authors provide more detail about this in the study selection section, but given that this is one of the main eligibility criteria, more clarity about exactly what would and would not count as a "validated" measure of resilience is important. The authors list five measures in the results section on page 9, but it is unclear whether those 5 measures represented a-priori the only acceptable measures of resilience.</p> <p>*The authors could be more explicit in listing the details of moderator analyses that were planned a-priori (pg 7)</p> <p>*More information could be provided on page 9 about the characteristics of interventions that were used to classify the treatment approach. There can be significant overlap in CBT and mindfulness interventions so the rationale for these subgroup analyses should be very clear. The extent to which these classifications are robust would impact whether/how it would be best to present those subgroup analyses.</p>
-------------------------	---

	<p>*In figure 1, were the 2 studies excluded due to poor quality and the 2 studies excluded due to poor design the same 2 studies. The separate column for the WHO trials makes it somewhat confusing about when those studies entered the flow of study selection</p> <p>*Standardized is misspelled in Supplementary Figure 1</p> <p>*It would be helpful to include a brief note in supplementary Table 1 to explain how Evidence of Effectiveness was operationalized. Right now the information in that column is generally just a Y or N.</p> <p>I think a meta-analytic review of the efficacy of resilience interventions could provide a useful contribution to the literature, but this manuscript would benefit from the inclusion of more details about the above issues.</p>
--	---

REVIEWER	Deborah Caldwell University of Bristol, UK
REVIEW RETURNED	07-Aug-2017

GENERAL COMMENTS	<p>This paper reports a systematic review of resilience training interventions. It is not clear, however, whether they are resilience interventions to increase resilience or resilience interventions to reduce mental ill-health. The introduction does not clarify this, nor do the aims reported on page 26, page 5 of the manuscript.</p> <p>Parts of the introduction are speculative and lack adequate references - e.g. "resilience would appear to go beyond simple genetic influence" and "more recently there has been a tendency to confuse resilience with well-being". Neither statement is supported by a reference. On the whole, the introduction would benefit from being shortened and kept factual.</p> <p>Page 5, line 36 to 41: "the main aim of the review is to synthesize available research evidence on resilience interventions..." Here the authors should state what the purpose of the interventions are - e.g. "interventions to promote/ increase resilience"? or "resilience interventions to prevent mental ill health/ increase coping skills" etc</p> <p>In the aims the authors state " we aim to only include studies that utilized valid and reliable measures of resilience as previously defined by Windle". However, they do not just include those defined by Windle, but end up allowing further scales to be included (presumably they were valid scales missed by the Windle review?). This brings me to a second point - I couldn't find a published protocol for this review online and the authors do not discuss there being one. Protocols for systematic reviews are essential, as they are for primary studies. I cannot judge the objectivity and transparency of much of the review without knowing what the authors originally intended.</p> <p>Page 5, line 47: " a systematic search was carried out according to PRISMA guidelines". This sentence is slightly misleading - PRISMA state that a systematic search should be carried out, but this reads as though PRISMA give guidance on systematic searching (a very minor point!). There is also a typo in this section "trails" instead of "trials". (line 54).</p> <p>Page 6, line 17. It is fine to include both RCTs and controlled trials in the review, however I cannot tell whether they were combined in a single analysis. These studies should be kept separate in any meta-</p>
-------------------------	--

	<p>analysis, or distinguished using subgroup analyses (using the group command in Stata, for example). The authors should also note that their eligibility criteria could introduce bias (e.g. exclusion of non-English publications) and that this is not addressed by the review. This is a considerable flaw. (note this about the methodological rigor of the review process and not the individual studies). Again, without a published protocol it is difficult to judge. The lack of access to a protocol is a serious flaw for this review - it should not have been marked "NA" in the PRISMA checklist, a protocol is applicable.</p> <p>Page 6, line 22: do the authors consider the potential for heterogeneity that allowing any comparator and any length of follow up might induce? Is this explored in subgroups analyses, or are there too few studies. There isn't enough information provided for me to judge.</p> <p>Page 6, line 36: The last sentence of the Study Selection paragraph contradicts page 5, line 40. Either the authors only included those scales identified by Windle, or they included scales they as a team deemed to be valid/ reliable etc. Or perhaps they did both - it is not clear and it needs to be consistent throughout the paper. Without access to a protocol I can't judge whether this was a post-hoc decision, a modification to what they intended or just a confusing write-up.</p> <p>Page 6, line 41 - Why did the authors use Downs and Black? Why did they modify it? One of the 3 references they provide is a self-citation and all are systematic reviews and not methodological papers which describe the need for the modification of the checklist. Again, a protocol would have provided the answer.</p> <p>Page 7, lines 9 to 34. The data synthesis section needs attention. The sentence "a positive effect indicates...a superior effect", should be re-written as it is just referring to the effect estimate and other factors should be taken into account when interpreting the pooled estimate - such as the 95% CIs, a p-value and the effect size. the degree of heterogeneity and number of studies are also important to consider. The authors state that studies were weighted by inverse variation method (I think they mean inverse-variance) and then go on to say they used a random effects model. As they used Stata, I assume they used the DerSimonian and Laird approach? Conventionally, the inverse-variance approach is a fixed effect model, although D&L is based on this. This should all be clarified in the text. The description of the random effects model is not quite accurate either and neither is the I^2 description -the whole section needs re-working. Note: ranges for the I^2 are not appropriate. The authors should also explicitly state all subgroup analyses they had planned.</p> <p>Page 9: results The authors state that there were 11 RCTs included, but from their list of study characteristics I note that 2 studies were controlled trials. Is this correct?</p> <p>Page 9, line 24: two of the scales were not in the earlier list of validated scales. It is not clear whether these were included to increase the number of studies available, or whether they are indeed validated scales.</p> <p>There is no detail on what the comparator interventions were. This is</p>
--	---

	<p>important as it could be a source of heterogeneity but also just of interest to the reader.</p> <p>Page 10, 14: which studies were "good quality", n=5 is insufficient information.</p> <p>The analyses which are subgroup analyses should be more clearly labeled. There is no discussion of heterogeneity, either in the text or in terms of I² metrics. In Figure 2 the Stata plot notes I² is 47% - what does this mean for the results? Why is I² missing from the 3 plots in Figure 3?</p> <p>Discussion, page 13. The conclusions are overstated: "resilience interventions... have a positive impact on resilience. This finding has far-reaching implications...". Again, as in the introduction, the authors make unsubstantiated claims without supporting references "Our findings highlight the benefits of mindfulness training...". There are only 2 studies which looked solely at mindfulness and 5 which looked at mixed interventions. This is simply an over-statement of findings.</p>
--	---

REVIEWER	Samprit Banerjee Weill Cornell Medical College, USA
REVIEW RETURNED	24-Aug-2017

GENERAL COMMENTS	<p>The authors conduct a meta-analysis of the effects of any intervention to improve resilience. The paper is well-written, the methods and results are clearly stated and the analysis has been conducted thoroughly. However, one major concern with this manuscript is the following -</p> <p>This study considers a wide-variety of interventions, active and passive controls, variable follow-up times, three measures of resilience and studies of different designs (randomised and non-randomised studies). With only 11 studies and so much heterogeneity I am not sure how to interpret the effect sizes estimated in this meta-analysis.</p> <p>Some minor comments -</p> <p>1) I do not agree with the following statement in the discussion - "There is less evidence regarding the long-term effect of resilience training but the research evidence thus far suggests that the positive impact of Mindfulness or CBT-based resilience training lasts up to 6-months." - This conclusion was not reached in this study and it is unclear how this statement was made.</p> <p>2) The time-frame of the studies included in not explicitly included in the eligibilty criteria but the time range of studies under consideration should be reported.</p> <p>3) The numbers in the Results section (17 studies) and that in the consort chart do not agree. Also, Figure 1 should show an additional step which reduces the number of studies to 11 which is the true number of studies used in the analysis.</p> <p>4) "Two practising psychologists reviewed the interventions.." - if they are authors of this study their initials should be indicated.</p> <p>5) I² for heterogeneity should be reported in Figure 3 for the sub-analyses.</p>
-------------------------	--

	<p>6) Supplementary Table 1 included more than 11 studies and is misleading. Only the studies that featured in the meta-analysis should be reported.</p> <p>7) Sensitivity analysis should be performed with respect to the three scales used to measure resilience (because different scales can differently weigh different dimensions of resilience) and type of study (RCT, RCT pilot and CT).</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 (Matthew Gallagher)

The authors appropriately follow the PRISMA guidelines in reporting the findings of their meta-analysis and the methods of the literature search and statistical analyses generally appear to be appropriate and conducted properly. There were some areas, however, where it would be beneficial to be more explicit about various methodological details of the meta-analysis. Specifically:

- It would be helpful to be more explicit in the methods section on page about was considered a validated measure of resilience and therefore met eligibility criteria for this meta-analysis. As the authors note, the definition and measurement of resilience is very complicated, so being very clear about what was considered unacceptable resilience outcome is crucial for understanding the scope and focus of this review. The authors provide more detail about this in the study selection section, but given that this is one of the main eligibility criteria, more clarity about exactly what would and would not count as a "validated" measure of resilience is important. The authors list five measures in the results section on page 9, but it is unclear whether those 5 measures represented a-priori the only acceptable measures of resilience.

We thank Reviewer 1 for raising this important point. We agree that we should have been clearer about this point. On page 6, we have now provided much more detail about how we assessed the measures of resilience. In this section, the following has now been added:

“Studies had to describe a specific aim to improve resilience and employ an acceptable measure of resilience as one of the outcome measures. It was decided a priori that in order to be an acceptable measure of resilience, the outcome measure used had to meet two criteria. Firstly, the measure had to assess an individual’s ability to adapt to change and cope effectively with significant life adversity. Secondly, the measure had to have undergone some type of validity assessment. As noted above, there is currently no gold standard measure of resilience (32), which makes assessing criterion validity difficult. In keeping with the conclusions by Windle, Bennet and Noyes (32) the following three measures of resilience were agreed a priori to have met both of the defined criteria; The Connors and Davidson Resilience Scale, The Brief Resilience Scale and The 14-item Resilience Scale. If studies were identified that employed other measures of resilience, these were closely examined in terms of the construct that was measured and the degree to which it had been validated against other recognized outcomes.”

- The authors could be more explicit in listing the details of moderator analyses that were planned a-priori (pg 7)

We apologise for this oversight. On page 7 we have now confirmed that there were two sub-analyses planned a priori; one examining the different types of interventions and another different lengths of follow up times.

- More information could be provided on page 9 about the characteristics of interventions that were used to classify the treatment approach. There can be significant overlap in CBT and

mindfulness interventions so the rationale for these subgroup analyses should be very clear. The extent to which these classifications are robust would impact whether/how it would be best to present those subgroup analyses.

We agree with this suggestion and apologise for this oversight. An additional paragraph with this information has now been added to page 7.

- In figure 1, were the 2 studies excluded due to poor quality and the 2 studies excluded due to poor design the same 2 studies. The separate column for the WHO trials makes it somewhat confusing about when those studies entered the flow of study selection

We agree that the separation of the WHO trials makes the flow diagram in Figure 1 somewhat confusing. We have now joined the trial registry data and the standard search data in the diagram, which hopefully helps make the numbers excluded easier to interpret.

- Standardized is misspelled in Supplementary Figure 1

This has been corrected

- It would be helpful to include a brief note in supplementary Table 1 to explain how Evidence of Effectiveness was operationalized. Right now the information in that column is generally just a Y or N.

We thank Reviewer 1 for this helpful suggestion. A foot note has been added to Supplementary Table 1 to explain this.

Reviewer: 2 (Deborah Caldwell)

This paper reports a systematic review of resilience training interventions. It is not clear, however, whether they are resilience interventions to increase resilience or resilience interventions to reduce mental ill-health. The introduction does not clarify this, nor do the aims reported on page 26, page 5 of the manuscript.

We apologise for this lack of clarity. We have now ensured that both the abstract and the introduction make it clear that the focus of this review was interventions to increase personal resilience.

Parts of the introduction are speculative and lack adequate references - e.g. "resilience would appear to go beyond simple genetic influence" and "more recently there has been a tendency to confuse resilience with well-being". Neither statement is supported by a reference. On the whole, the introduction would benefit from being shortened and kept factual.

We thank Reviewer 2 for these suggestions. We have removed the two statements highlighted, ensured that all other statements are adequately referenced and factual and have shortened the overall length of the introduction.

Page 5, line 36 to 41: "the main aim of the review is to synthesize available research evidence on resilience interventions..." Here the authors should state what the purpose of the interventions are - e.g. "interventions to promote/ increase resilience"? or "resilience interventions to prevent mental ill health/ increase coping skills" etc

This sentence has been reworded to make it clear that our aim was to examine interventions designed to promote or enhance individual resilience.

In the aims the authors state " we aim to only include studies that utilized valid and reliable measures of resilience as previously defined by Windle". However, they do not just include those defined by Windle, but end up allowing further scales to be included (presumably they were valid scales missed by the Windle review?).

As outlined in our response to Reviewer 1, we have now expanded the section of the methods (on page 5) which describes our a priori decisions about what would constitute an acceptable measure of resilience. In short, we defined two criteria that needed to be met; firstly, the measure had to assess an individual's ability to adapt to change and cope effectively with significant life adversity. Secondly, the measure had to have undergone some type of validity assessment. We used the results of Windle's review to assist with this, but have also now added an explanation of how measures not included in their review were assessed.

This brings me to a second point - I couldn't find a published protocol for this review online and the authors do not discuss there being one. Protocols for systematic reviews are essential, as they are for primary studies. I cannot judge the objectivity and transparency of much of the review without knowing what the authors originally intended.

Reviewer 2 is correct that we have not published a protocol for this review. We are aware that bodies such as the Cochrane Collaboration have been promoting published protocols for systematic reviews for a number of years, but we would respectfully observe that the majority of published systematic reviews do not have a published protocol. We also note that the BMJ Open's Editorial Policies only demand prospective protocol registration for clinical trials. In modifying our manuscript we have attempted to make clear when decisions were made a priori. We also take on board Reviewer 2's point regarding the value of published protocols for systematic reviews and will certainly aim to publish prospective protocols for future reviews we undertake.

Page 5, line 47: " a systematic search was carried out according to PRISMA guidelines". This sentence is slightly misleading - PRISMA state that a systematic search should be carried out, but this reads as though PRISMA give guidance on systematic searching (a very minor point!). There is also a typo in this section "trails" instead of "trials". (line 54).

We thank Reviewer 2 for highlighting these two errors, which have now been corrected.

Page 6, line 17. It is fine to include both RCTs and controlled trials in the review, however I cannot tell whether they were combined in a single analysis. These studies should be kept separate in any meta-analysis, or distinguished using subgroup analyses (using the group command in Stata, for example).

We absolutely agree with Reviewer 2 on this point. While we looked for both RCTs and controlled trials in our search strategy, all eleven of the studies that ended up contributing to the meta-analysis were RCTs. We have now made this point clearer.

The authors should also note that their eligibility criteria could introduce bias (e.g. exclusion of non-English publications) and that this is not addressed by the review. This is a considerable flaw. (note this about the methodological rigor of the review process and not the individual studies). Again, without a published protocol it is difficult to judge. The lack of access to a protocol is a serious flaw for this review - it should not have been marked "NA" in the PRISMA checklist, a protocol is applicable.

Each of these issues have now been added to our discussion about the limitations of our review on page 13.

Page 6, line 22: do the authors consider the potential for heterogeneity that allowing any comparator and any length of follow up might induce? Is this explored in subgroups analyses, or are there too few studies. There isn't enough information provided for me to judge.

We agree that considering possible causes of heterogeneity is an important part of any meta-analysis. However, as suggested by Reviewer 2, the limited number of studies made this very difficult. While we were able to stratify our analysis according to the program type (CBT and/or mindfulness), there were only two studies that utilised a control condition other than waitlist, making it impossible to examine the impact of this factor. We have added a discussion about this on pages 13 and 14.

Page 6, line 36: The last sentence of the Study Selection paragraph contradicts page 5, line 40. Either the authors only included those scales identified by Windle, or they included scales they as a team deemed to be valid/ reliable etc. Or perhaps they did both - it is not clear and it needs to be consistent throughout the paper. Without access to a protocol I can't judge whether this was a post-hoc decision, a modification to what they intended or just a confusing write-up.

We apologise for this lack of clarity. We have attempted to make our approach much clearer, with a detailed explanation in the methods section on page 5.

Page 6, line 41 - Why did the authors use Downs and Black? Why did they modify it? One of the 3 references they provide is a self-citation and all are systematic reviews and not methodological papers which describe the need for the modification of the checklist. Again, a protocol would have provided the answer.

The Downs and Black checklist was chosen because it was one of the two tools for assessing methodological quality of intervention studies cited in the Cochrane Handbook of Systematic Reviews. There appears to have been a problem with one of the references in our original submission. The references were meant to be previous published systematic reviews which had used this minor modification. We apologise for this error. This has now been corrected.

Page 7, lines 9 to 34. The data synthesis section needs attention. The sentence "a positive effect indicates...a superior effect", should be re-written as it is just referring to the effect estimate and other factors should be taken into account when interpreting the pooled estimate - such as the 95% CIs, a p-value and the effect size. the degree of heterogeneity and number of studies are also important to consider. The authors state that studies were weighted by inverse variation method (I think they mean inverse-variance) and then go on to say they used a random effects model. As they used Stata, I assume they used the DerSimonian and Laird approach? Conventionally, the inverse-variance approach is a fixed effect model, although D&L is based on this. This should all be clarified in the text. The description of the random effects model is not quite accurate either and neither is the I^2 description -the whole section needs re-working. Note: ranges for the I^2 are not appropriate. The authors should also explicitly state all subgroup analyses they had planned.

As suggested by Reviewer 2, this section has been re-written. The statements questioned by Reviewer 2 have been removed, we have clarified that the DerSimonian and Laird approach was used and we have explicitly stated the two subgroup analyses that were planned a priori. The changes are on pages 6 and 7 of the modified manuscript.

Page 9: results

The authors state that there were 11 RCTs included, but from their list of study characteristics I note that 2 studies were controlled trials. Is this correct?

We apologise for this lack of clarity. As discussed above, all of the 11 trials included in the meta-analysis were RCTs. This has now been made clearer in the Results section on page 9.

Page 9, line 24: two of the scales were not in the earlier list of validated scales. It is not clear whether these were included to increase the number of studies available, or whether they are indeed validated scales.

As outlined above, we have now added further details about the resilience scales and the assessment process undertaken on page 6 of the modified manuscript.

There is no detail on what the comparator interventions were. This is important as it could be a source of heterogeneity but also just of interest to the reader. Page 10, 14: which studies were "good quality", n=5 is insufficient information.

This information was provided in Supplementary Table 1. In all but two of the studies included in the meta-analysis the control condition was a waitlist. As noted above, a discussion about the potential role that the various control conditions may have had on our results is now included on pages 13 and 14. We have also directed the reader to the information contained in the supplementary table at the appropriate points of the Results section on page 9.

The analyses which are subgroup analyses should be more clearly labelled. There is no discussion of heterogeneity, either in the text or in terms of I^2 metrics. In Figure 2 the Stata plot notes I^2 is 47% - what does this mean for the results?

As suggested, subgroup analyses have now been clearly labelled. In addition, a more detailed discussion of the heterogeneity and I^2 metric have now been included in both the Results section and in the Discussion (pages 13 and 14)

Discussion, page 13. The conclusions are overstated: "resilience interventions... have a positive impact on resilience. This finding has far-reaching implications...". Again, as in the introduction, the authors make unsubstantiated claims without supporting references "Our findings highlight the benefits of mindfulness training...". There are only 2 studies which looked solely at mindfulness and 5 which looked at mixed interventions. This is simply an over-statement of findings.

On reflection we agree with Reviewer 2 regarding these conclusions. We have re-written these parts of the Discussion to ensure that the statements made reflect our results and the limitations outlined above. The specific statements noted by Reviewer 2 have been removed.

Reviewer: 3 (Samprit Banerjee)

The authors conduct a meta-analysis of the effects of any intervention to improve resilience. The paper is well-written, the methods and results are clearly stated and the analysis has been conducted thoroughly. However, one major concern with this manuscript is the following - This study considers a wide-variety of interventions, active and passive controls, variable follow-up times, three measures of resilience and studies of different designs (randomised and non-randomised studies). With only 11 studies and so much heterogeneity I am not sure how to interpret the effect sizes estimated in this meta-analysis.

We thank Reviewer 3 for their kind comments. We agree that the issue of heterogeneity is very important and deserved greater consideration in our manuscript. As outlined in our detailed responses to Reviewer 1 and 2, we have now made a number of changes to our manuscript which we feel will address this concern. In particular, we have been clearer that all 11 studies included in the meta-

analysis were randomised controlled trials and that all but two utilised the same control condition (wait list). We have also added a detailed discussion about heterogeneity and the limitations this places on interpretation in the Discussion on pages 13 and 14.

Some minor comments -

1) I do not agree with the following statement in the discussion - "There is less evidence regarding the long-term effect of resilience training but the research evidence thus far suggests that the positive impact of Mindfulness or CBT-based resilience training lasts up to 6-months." - This conclusion was not reached in this study and it is unclear how this statement was made.

This statement has now been removed from the Discussion

2) The time-frame of the studies included in not explicitly included in the eligibility criteria but the time range of studies under consideration should be reported.

We apologise for this lack of clarity. On page 5, under the sub-heading 'Search Strategy' we now make it clear that no time restrictions were placed on the search strategy, with all published articles up to June 2016 considered eligible.

3) The numbers in the Results section (17 studies) and that in the consort chart do not agree. Also, Figure 1 should show an additional step which reduces the number of studies to 11 which is the true number of studies used in the analysis.

As noted in our response to Reviewer 1, the flow diagram in Figure 1 has now been modified, which we hope addresses this point.

4) "Two practicing psychologists reviewed the interventions.." - if they are authors of this study their initials should be indicated.

We apologise for this oversight, this has now been corrected on page 6.

5) I² for heterogeneity should be reported in Figure 3 for the sub-analyses.

Done

6) Supplementary Table 1 included more than 11 studies and is misleading. Only the studies that featured in the meta-analysis should be reported.

We agree with this point and have modified the table accordingly.

7) Sensitivity analysis should be performed with respect to the three scales used to measure resilience (because different scales can differently weigh different dimensions of resilience) and type of study (RCT, RCT pilot and CT).

As noted in our responses to the other reviewers above, all 11 studies included in the meta-analyses were RCTs. We have now made this clearer in our Results section. We have also included in our Discussion a commentary on how sub-analyses of some variables, such as the specific resilience scale used or control condition, were not possible due to the relatively small number of studies within each subgroup. The limitations that this places upon our conclusions are noted on pages 13 and 14 and the overall tone of our Discussion has been modified accordingly.

Comments from the Editorial Team

More details about the study's methods is needed in the abstract (for example, what data sources were used in the literature search? What was the eligibility criteria?). We recommend taking a look at the abstracts of other systematic reviews published in BMJ Open as examples.

We thank the editorial team for this advice. We have looked at a number of other systematic reviews previously published in BMJ Open and have adapted and added to our abstract accordingly.

Please provide the full search strategy for at least one database as a supplementary file and refer to this in the methods section.

Done

Please justify your inclusion criteria. Your PICO seems too vague. Why was it not restricted to healthcare professionals for example? Studies include not only doctors, but also patients and members of the armed forces.

We are a bit confused by this comment as our search was for trials involving any group of adults. There was never any intention to limit our search to health professionals or any other occupational group. However, as outlined below, we have now included a detailed discussion of how these results would apply to health professionals.

We are not really told anything about the data extraction process (what data was extracted, how many authors extracted the data, etc). Please add this to the paper.

Information about this has been added on page 6, under a new subheading 'Data extraction and contact with researchers'.

Of the 111 papers you reviewed in full, what were the reasons of exclusion afterwards? This is not clear.

As outlined in our responses to the reviewers comments above, we have now redrawn figure 1 and redrafted the first paragraph of the Results. We have now included a clear statement that 96 of the 111 papers were excluded as they did not meet the inclusion criteria.

You do not comment on heterogeneity, and it is not reported in the forest plots. However, these are studies in very diverse populations and employing very different interventions. This needs to be considered in the paper.

As outlined above, we agree that this was an oversight in our original manuscript. We have now made sure to add the heterogeneity estimates in the Results section and have included a detailed discussions of the implications resulting from this on pages 13 and 14 of the Discussion.

The forest plots do not seem to be well reported. Please check these. We recommend taking a look at the following guidance: <https://www.cebi.ox.ac.uk/for-practitioners/what-is-good-evidence/how-to-read-a-forest-plot.html>

We thank the editorial team for providing this useful link. We have read it in detail and tried to modify our reporting of the main meta-analysis to match the two-step approach outlined.

The findings seem very difficult to interpret because of the heterogeneity, the wide confidence intervals (some ranging from a very small to a very large effect size) and the use of difference resilience scales. As such we recommend toning down the importance of your findings and conclusions. Your discussion seems quite generic, and we would like you to comment more on the relevance of the findings to doctors. You should perhaps also cite and discuss this recently published paper:

Nurse Educ Today. 2017 Nov;58:65-71. doi: 10.1016/j.nedt.2017.07.018. Epub 2017 Aug 16.
What do we know about student resilience in health professional education? A scoping review of the literature.
<https://www.ncbi.nlm.nih.gov/pubmed/28898767>

As outlined above, we agree that the tone of our Discussion needed to be toned down. We have reviewed the entire Discussion and re-written parts to conform with this suggestion. We have also added a new paragraph to the Discussion which considers the relevance of our findings to the medical workforce. The suggested reference has been added to this part of the Discussion.