## 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)** | Trends and Predictors of Biomedical Research Quality, 1990-2015: A Meta-Research Study |
| **AUTHORS** | Catillon, Maryaline |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Paul Glasziou<br>Bond University<br>Co-Director, Australian EQUATOR Centre. |
| **REVIEW RETURNED** | 27-Mar-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a useful review, which looked at the trends in quality in over 20,000 trials assessed for 6 elements of Risk of Bias in over 2,000 Cochrane Reviews. The results apparently show improvement over time, particularly in the reporting of some elements. The progress has been slow, but seen over the 25 years in Figure 2, the apparent steady improvement is heartening. Some limitations not mentioned in "Limitations" are that Cochrane reviewers may have been able to obtain more information on more recent studies (from authors, registries, or protocols rather than the primary report) - that would soften the conclusion to one of improvement in access to study details.<br>A few other comments on presentation.<br>1. The authors refer to "biomedical research articles" and "studies" but aren't virtually all of these controlled trials? I would prefer that term, as otherwise readers may generalize to other types of research.<br>2. I could not match the numbers in the Figure A1 flow chart with other figures. For example,<br>Figure 1 is N=20,571 but in Fig1A that is Sample 1 (time series)<br>Figure 2 has multiple N's that I could not match up, but I thought was the time series?<br>3. Figure 1 might be easier to read as a stacked bar chart, and include the absolute numbers inside each bar segment (but keep as a % overall).<br>4. Figure 2. A key summary figure, but 4 of 7 use 0-80% and others use 70, 60, and 60%. I would suggest forcing 80% on all to make comparing each simpler.<br>5. Figure 3. Graph should indicate (i) Relative Risk as the measure, and (ii) on each side of 1 indicate poorer methods/better methods; poorer reporting/better reporting. I also wondered if it would be easier to read with the reference category included rather than in the footnote. |

| | |
|---|---|
| | 6. I would suggest a more extensive Discussion of the related literature on the changes of reporting due to CONSORT and other inititiaves, eg<br>Turner L1, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev. 2012 Nov 29;1:60. doi: 10.1186/2046-4053-1-60.<br>Moseley AM1, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. J Clin Epidemiol. 2011 Jun;64(6):594-601. doi: 10.1016/j.jclinepi.2010.08.009. Epub 2010 Dec 8.<br>Saltaji H1, Armijo-Olivo S2, Cummings GG3, Amin M1, Flores-Mir C1. Randomized clinical trials in dentistry: Risks of bias, risks of random errors, reporting quality, and methodologic quality over the years 1955-2013. PLoS One. 2017 Dec 22;12(12):e0190089. doi: 10.1371/journal.pone.0190089. eCollection 2017. |

| | |
|---|---|
| **REVIEWER** | Simon Gandevia<br>NeuRA (Neuroscience Research Australia), Australia |
| **REVIEW RETURNED** | 15-Apr-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is an extensive analysis of the quality of medical research using sophisticated meta-research methods. It paints strongly a worrying picture of the landscape. Many will not be surprised by this, as it has been pointed to by smaller studies, but the magnitude of the effect may come as a wake up call to many. In sequence through the manuscript I have the following issues:<br><br>• Page 2, paragraph beginning “Data”, Line 3: Suggest adding “was” before “mapped to bibliometric and funding data”.<br>• Page 3, dot point beginning “PubMed identifier”: Please give the percentage of studies for which the data were not available.<br>• Page 5: Sample: paragraph 2: It is stated that study quality was assessed consistently in virtually all the studies. Please indicate how this reliability of the assessment was determined. It is a weakness of the current submission that some methodological points are not explicitly covered.<br>• Page 5 analysis: exactly how was this done and what was the reliability? Was only one assessor used? What checks were in place?<br>• Page 6, paragraph 2: I do not wish to weaken the major message of the study. However, methods were said to be “inadequate” if bias was high in one or more reasons. Why was the threshold of one selected? Some might argue that it sets a very high bar. I am not suggesting that it is altered, but it might be helpful to indicate the general changes in the results if that one was changed to a higher number, say two. This point extends to a consideration of the definition of “unclear” risk of bias. Again, the threshold is a single dimension.<br>• Page 7, line 8: Perhaps it would be helpful to explain what is meant by the term “procedure”. Presumably it means more than just a surgical technique.<br>• Page 7: The description of table 1 could perhaps indicate the threshold that was used for a consideration of whether methods were adequate or not. |

| | |
|---|---|
| | • Page 8: Final paragraph: Please explain what is meant by the term “evolutions”. Presumably it means temporal pattern.<br>• Page 10, paragraph 2: The observation about first author affiliations with the pharmaceutical industry is interesting. How were multiple affiliations dealt with? In addition, did you consider looking at the affiliation of the corresponding author? My thought is that this may give a more interesting link to different universities.<br>• Table 1: It may help the reader if there is a vertical line placed between the column labelled “All” and “Adequate”.<br>• In Figure 2: the legend may be helped by adding a reference to Table A1.<br>• In Figure 3: it may be helpful to explain the horizontal arrow heads on several of the relative risk intervals.<br>• In Figure A1: I am not clear what is meant by the words in the box “and matching each study with its main reference.”<br>• In Figure A2: again, this seems to show a reasonably linear improvement in the prevalence of papers with a low risk of bias. Perhaps this pattern could be mentioned in the text.<br>• In Table A2: I may not have followed the statistical methodology correctly. I presume that the prediction model is relative to the category of “adequate methods”? This table suggests two other points which could be mentioned: the effect of the number of authors is possibly rather small, although its likelihood of statistical significance is high. If this is correct, perhaps it should be noted. The other perhaps surprising finding is the relative risk ratio for the use of a device. Is a comment warranted on this?<br>• Table A3: It may be helpful to the reader to have more detail in the final line beginning “Regression results predicting …”. |

## VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1 Paul Glasziou

This is a useful review, which looked at the trends in quality in over 20,000 trials assessed for 6 elements of Risk of Bias in over 2,000 Cochrane Reviews. The results apparently show improvement over time, particularly in the reporting of some elements. The progress has been slow, but seen over the 25 years in Figure 2 (renamed figure 3), the apparent steady improvement is heartening.

Some limitations not mentioned in "Limitations" are that Cochrane reviewers may have been able to obtain more information on more recent studies (from authors, registries, or protocols rather than the primary report) - that would soften the conclusion to one of improvement in access to study details.

Thank you for this important insight. I added the sentence below to the limitations:

“Cochrane reviewers may have been able to obtain more information on more recent RCTs (from authors, registries, or protocols rather than the primary report), suggesting that some of the apparent improvement in reporting may in fact be an improvement in access to study details.”

A few other comments on presentation.

1. The authors refer to "biomedical research articles" and "studies" but aren't virtually all of these controlled trials? I would prefer that term, as otherwise readers may generalize to other types of research.

I standardized the vocabulary, and now use “randomized controlled trial” or “RCT” consistently throughout the paper.

2. I could not match the numbers in the Figure a1 (renamed figure 1) flow chart with other figures. For example,

Figure 1 (renamed figure 2) is N=20,571 but in Fig1A that is Sample 1 (time series) Figure 2 (renamed figure 3) has multiple N's that I could not match up, but I thought was the time series?

I provide a new Figure a1 (renamed figure 1) for the data flow and specifically indicate which sample is used in each figure and table.

The last two boxes of the revised Figure a1 (renamed figure 1) now include:

20,571 RCTs, Main sample: used in Table 1, Figure 1 (renamed figure 2) and Figure 2 (renamed figure 3).

11,686 RCTs: Sample used in regressions including funding: Figure 3 (renamed figure 5), Table A2 and Table A3.

3. Figure 1 (renamed figure 2) might be easier to read as a stacked bar chart, and include the absolute numbers inside each bar segment (but keep as a % overall).

Thank you very much for this point. I provide a new Figure 1 (renamed figure 2) that is much easier to read and include the absolute numbers as well as the percentages of RCTs in each category.

4. Figure 2 (renamed figure 3). A key summary figure, but 4 of 7 use 0-80% and others use 70, 60, and 60%. I would suggest forcing 80% on all to make comparing each simpler.

I provide a new Figure 2 (renamed figure 3) forcing all y scales to be 0-80% to make comparisons simpler.

5. Figure 3 (renamed figure 5). Graph should indicate (i) Relative Risk as the measure, and (ii) on each side of 1 indicate poorer methods/better methods; poorer reporting/better reporting. I also wondered if it would be easier to read with the reference category included rather than in the footnote.

My revised new Figure 3 (renamed figure 5) includes these elements.

6. I would suggest a more extensive Discussion of the related literature on the changes of reporting due to CONSORT and other inititiaves, eg Turner L1, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev.

2012 Nov 29;1:60. doi: 10.1186/2046-4053-1-60.

Moseley AM1, Herbert RD, Maher CG, Sherrington C, Elkins MR. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. J Clin Epidemiol. 2011 Jun;64(6):594-601. doi: 10.1016/j.jclinepi.2010.08.009.

Epub 2010 Dec 8.

Saltaji H1, Armijo-Olivo S2, Cummings GG3, Amin M1, Flores-Mir C1. Randomized clinical trials in dentistry: Risks of bias, risks of random errors, reporting quality, and methodologic quality over the years 1955-2013. PLoS One. 2017 Dec 22;12(12):e0190089. doi: 10.1371/journal.pone.0190089. eCollection 2017.

Thank you for noting this. I added the following paragraphs to the discussion:

The overall proportion of poorly reported trials decreased by about 5 percentage points per decade. This is good news but much remains to be done. At the current rate of improvement, it would take 50 years for 95% of RCTs to be adequately reported. These results are consistent with previous research finding improvements in reporting in several clinical areas such as physiotherapy[47], and dentistry[25]. The trends for each dimension assessed separately are also very similar to those found in another large sample of RCTs.[24]

These improvement in reporting happened over a period of time when the Consolidated Standards of Reporting Trials (CONSORT) statement, a minimum set of evidencebased reporting recommendations, and other initiatives, such as the EQUATOR Network, developed to improve reporting practices.[41-45]. Since the 1990s, the CONSORT statement has been endorsed by over 50% of the core clinical journals indexed in PubMed and may improve reporting of RCTs they publish.[46] Spurred by the CONSORT statement, the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network, was launched in 2008 in the UK to improve the reliability of medical publications by promoting transparent and accurate reporting of health research.[47] Since, it has developed into a global initiative aiming to improve research reporting worldwide.[36]

Reviewer: 2

Simon Gandevia

This is an extensive analysis of the quality of medical research using sophisticated meta-research methods.  It paints strongly a worrying picture of the landscape.  Many will not be surprised by this, as it has been pointed to by smaller studies, but the magnitude of the effect may come as a wake up call to many.  In sequence through the manuscript I have the following issues:

- Page 2, paragraph beginning “Data”, Line 3:  Suggest adding “was” before

“mapped to bibliometric and funding data”.

I changed the paragraph to read: “Risk of bias assessments for random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting, for each study, were mapped to bibliometric and funding data”.

- Page 3, dot point beginning “PubMed identifier”:  Please give the percentage of studies for which the data were not available.

I added the percentage of studies for which data were not available:

“PubMed identifier, full-text and/or funding information were not available for all RCTs. 30.5% of RCTs (unpublished or published in journals not indexed in PubMed) did not have a PubMed identifier. 43.2% of RCTs with PubMed identifier did not have a full-text available from the Harvard Library. 23.6% of included RCTs were reported in articles disclosing NIH or industry funding.”

- Page 5:  Sample:  paragraph 2:  It is stated that study quality was assessed consistently in virtually all the studies.  Please indicate how this reliability of the assessment was determined. It is a weakness of the current submission that some methodological points are not explicitly covered.

This is very helpful.

I replaced the sentence about consistent assessments by the following sentence, expressing more clearly what “assessed consistently” meant, and including an example:

“Duplicates were removed. RCTs assessed multiple times with different outcomes (e.g., high risk in one review, unclear risk in another), were dropped.”

To address the more general point about methodology, I provided more detailed captions for the tables and figures and provided more details about data and methods (see following point).

- Page 5 analysis: exactly how was this done and what was the reliability? Was only one assessor used? What checks were in place?

Thank you! I extended the methods section to provide more detailed information:

Cochrane reviews constitute a valuable data source to assess biomedical research quality as they follow strict methods and precise reporting guidelines defined in the Cochrane Handbook.[29-30] This study does not involve new assessment of the methods and reporting of included RCTs, but relies entirely on the assessments available in the Cochrane reviews, which are systematically performed by two expert reviewers who compare their assessments and reach consensus on the final assessment.[29] The research method dimensions evaluated in Cochrane reviews include random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting (detailed in Supplementary Table A1).[31]

The database assembly had seven steps: (1) All included references were extracted from each review, including PubMed identifiers. (2) All risk of bias assessments on the six dimensions of the 2011 update of the Cochrane Risk of Bias Assessment Tool (see Table A1) were extracted from each review. Each assessment included three variables: bias type (e.g., random sequence generation), judgement (e.g., low risk) and support for judgement (e.g., computer random number generator). (3) Each RCT was matched with its main published reference as identified by Cochrane reviewers. (4) PubMed records corresponding to these publications, including bibliometric information and first author affiliation, were retrieved using the E-utilities public API. (5) Affiliation information for other authors (not available from PubMed over the study period) was retrieved from SCOPUS. (6) Full-text for references with PubMed identifier were retrieved from the Harvard Library. (7) Industry funding information was extracted from the full-text PDFs.

- Page 6, paragraph 2: I do not wish to weaken the major message of the study. However, methods were said to be “inadequate” if bias was high in one or more reasons. Why was the threshold of one selected? Some might argue that it sets a very high bar. I am not suggesting that it is altered, but it might be helpful to indicate the general changes in the results if that one was changed to a higher number, say two. This point extends to a consideration of the definition of

“unclear” risk of bias. Again, the threshold is a single dimension.

This is a very good question. I added the following paragraph in the “Analysis” section:

“Following guidelines for assessing the quality of evidence[31] and previous empirical work [7], the RCT-level assessment was “adequate methods” if the study was at low risk of bias on all dimensions assessed. It was “inadequate methods” if the study was at high risk of bias for one or more reasons. It was “poorly reported” if the reviewers did not have enough information to assess whether the methods used were adequate or inadequate (if the study was at “unclear” risk of bias for at least one reason). Several reasons support the use of at least one high risk of bias assessment as the definition for inadequate methods. Some risk of bias domains might translate into more statistical bias than others, but empirical evidence on the relative importance of the risk of bias domains is limited, and the effect of several versus one high risk assessment on research outcomes is unknown[32,33] The empirical relationship between risk of bias assessments and research outcomes (including actual statistical bias) requires further research.

There is also a theoretical reason to use at least one high risk of bias assessment as the definition of method inadequacy. Cochrane risk of bias domains can be mapped to important conditions to make RCTs valuable. If not truly randomized or if differences between the treatment and control group are introduced post-randomization, a RCT may not produce an unbiased estimate of the treatment effect.[34] These two conditions imply that one inadequacy in the randomization process (non-random sequence generation or inadequate allocation concealment), or one difference introduced post randomization between the treatment and control groups (through inadequate blinding of participants, personnel, or outcome assessors) or after the trial (due to incomplete outcome data or selective reporting) should be the default threshold for assessing methods adequacy.”

- Page 7, line 8: Perhaps it would be helpful to explain what is meant by the term “procedure”. Presumably it means more than just a surgical technique.

Procedure meant “surgical procedure”, I changed it to “surgery” to clarify.

- Page 7: The description of table 1 could perhaps indicate the threshold that was used for a consideration of whether methods were adequate or not.

I modified the description of Table 1 in the following way:

“Table 1. Descriptive statistics. Unless otherwise specified, column 1 reports the number of RCTs and their proportion as of the total number of RCTs (N=20,571). A RCT uses adequate methods if it is at “low risk of bias” on all six dimensions assessed

(see Table A1). Methods are inadequate if a RCT is at “high risk of bias” for at least one reason. Methods are poorly reported if there is no evidence of methods inadequacy, but at least one assessment is “unclear risk of bias.””

- Page 8: Final paragraph: Please explain what is meant by the term

“evolutions”. Presumably it means temporal pattern.

Yes, I modified this paragraph to clarify this point:

“Similar patterns suggest that the evolution over time observed for the RCTs assessed on all dimensions (N=20,571) reflects the evolution over time in all RCTs assessed on at least one dimension.”

- Page 10, paragraph 2: The observation about first author affiliations with the pharmaceutical industry is interesting. How were multiple affiliations dealt with? In addition, did you consider looking at the affiliation of the corresponding author? My thought is that this may give a more interesting link to different universities.

This is an important point that I will consider in future work, but might require a different approach given current data availability. The database includes the primary affiliation sector listed in PubMed, complemented by affiliation from SCOPUS if not available in PubMed, and whether any other author was affiliated with industry. Classification of sectors relies on primary reported affiliation. Corresponding Author information is not available consistently in existing databases.

In the limitations I added: “Classification of sectors relies on primary reported affiliation.”

- Table 1: It may help the reader if there is a vertical line placed between the column labelled “All” and “Adequate”.

Thanks for noting this! I did and the new Table 1 reads better.

- In Figure 2 (renamed figure 3): the legend may be helped by adding a reference to Table A1.

I modified the legend of Figure 2 (renamed figure 3) in the following way:

“Fig.2. Evolution of methods and reporting over time. A.: Proportion of RCTs using adequate methods, inadequate methods and poorly reported. B. to G.: Proportion of RCTs at low risk of bias, high risk of bias and unclear risk of bias for each dimension assessed. See Table A1 for the definition of each dimension. N=20,571 RCTs. An observation is a RCT assessed on all six dimensions. See Figure a1 (renamed figure 1)for more detailed information about the sample."

- In Figure 3 (renamed figure 5): it may be helpful to explain the horizontal arrow heads on several of the relative risk intervals.

I added the following sentence in the legend of Figure 3 (renamed figure 5):

The arrow heads on the confidence intervals indicate that the upper bound of the 95% confidence interval is greater than 2.

- In Figure a1 (renamed figure 1) I am not clear what is meant by the words in the box "and matching each study with its main reference."

I provide a new Figure a1 (renamed figure 1), including a new formulation for this step:

"Matching RCTs reported in several publications indexed in PubMed to their main reference, as identified by Cochrane reviewers."

Cochrane reviewers give a name to each RCT, a "study name" corresponding to the main publication reporting the trial. For instance, if there are three publications about a trial, Cochrane reviewers will cite these three publications in the included references, but they will give to the RCT the "study name" corresponding to the main publication. The bibliometric information associated with each RCT is retrieved from this main publication.

- In Figure a2 (renamed figure 4): again, this seems to show a reasonably linear improvement in the prevalence of papers with a low risk of bias. Perhaps this pattern could be mentioned in the text.

Thank you for this observation.

I provide a new Figure a2 (renamed figure 4), including a linear trendline for the improvement over time in adequate methods.

I also modify the result section and the discussion section in the following ways:

In the “Results” section, I include the following sentence:

“The proportion of RCTs using adequate methods increased linearly, by 3 percentage points per decade, from 2.6% in 1990 to 10.3% in 2015.”

In the “Discussion” section, I include the following sentence:

“The linear increase in the proportion of RCTs using adequate methods is heartening. However, at the current rate of improvement (3 percentage points per decade), it would take more than a century for half of RCTs to use adequate methods.”

- In Table A2: I may not have followed the statistical methodology correctly. I presume that the prediction model is relative to the category of “adequate methods”? This table suggests two other points which could be mentioned: the effect of the number of authors is possibly rather small, although its likelihood of statistical significance is high. If this is correct, perhaps it should be noted. The other perhaps surprising finding is the relative risk ratio for the use of a device. Is a comment warranted on this?

I changed the caption of Table A2. In the discussion, I included a comment on number of authors in the paragraph on team characteristics and I added a new paragraph commenting on results for drugs versus devices.

The new caption of Table A2 is:

“This table presents main regression results (relative risk ratios and p-values) from estimating the multinomial logit model predicting overall RCT quality. The dependent variable is a categorical variable and can take three values: adequate methods, inadequate methods and poor reporting. Adequate methods is the reference outcome category. The regression sample includes 11,686 RCTs with accessible full-text. See Figure a1 (renamed figure 1)for more detailed information about the sample. The relative risk ratios represent the likelihood of a RCT with specific funding, sector, study/team, technology and country characteristics using inadequate methods (or being poorly reported), as compared to the likelihood of a RCT in a reference group without these characteristics using inadequate methods (or being poorly reported). In the regression, sector, technology and country are categorical variables. The omitted category for sector is other university. The omitted category for technology is other interventions. The omitted category for country is other countries. The regression includes topic and year fixed effects. These results are plotted in Figure 3 (renamed figure 5). Other regression results predicting relative risk ratios for high or unclear risk of bias on each dimension assessed (as opposed to overall quality) are reported in Table A3.

The new comments on number of authors reads:

“Increasing the number of authors by one was associated with a small, but highly significant improvement in methods and reporting. Many RCTs are published by large teams so it is not surprising that the effect of an additional author was small. But this effect was also highly significant, consistent with previous research finding that larger teams and international teams produce more frequently cited research, [41,42] (…)”

The new paragraph on drugs versus devices reads:

“Finally, RCTs on drugs were more likely to use adequate methods than RCTs on other interventions, while RCTs on devices were more likely to use inadequate methods. In many countries, trials on drugs are more tightly regulated than trials on devices. In the US, under the Federal Food, Drug, and Cosmetic Act (FDCA, 1938), drugs and devices face different premarket review and post-market compliance requirements. The finding is also consistent with specific barriers to the conduct of RCTs on medical devices, in particular for randomization and blinding, and with the lack of scientific advice and regulations for medical device trials [60]. RCTs on drugs were using better methods and reporting than RCTs on other interventions, but much remains to be done. This finding is consistent with previous work showing that even RCTs used in the drug approval process frequently use inadequate methods and reporting.[61]”

- Table A3: It may be helpful to the reader to have more detail in the final line beginning “Regression results predicting …”.

I modified the legend of Table A3 in the following way:

“*p<0.05, **p<0.01, ***p<0.001

This table presents main regression results (relative risk ratios and p-values) from estimating the multinomial logit model predicting risk of bias for each dimension assessed. The dependent variable is a categorical variable and can take three values:

low risk, high risk or unclear risk. Low risk is the reference outcome category. The regression sample includes 11,686 RCTs with accessible full-text. See Figure a1 (renamed figure 1)for more detailed information about the sample. The relative risk ratios represent the likelihood of a RCT with specific funding, sector, study/team, technology and country characteristics being assessed at high risk (or unclear risk) as compared to the likelihood of a RCT in a reference group without these characteristics being assessed at high risk (or unclear risk). In the regression, sector, technology and country are categorical variables. The omitted category for sector is other university. The omitted category for technology is other interventions. The omitted category for country is other countries. The regression includes topic and year fixed effects. Other regression results predicting relative risk ratios for overall quality (as opposed to relative risk ratios for high or unclear risk of bias on each dimension assessed) are reported in Table A2.”

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Simon Gandevia<br>NeuRA (Neuroscience Research Australia), Australia |
| **REVIEW RETURNED** | 14-Jun-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | In my view the author has made major improvements to the manuscript in response to the two sets of reviewers' comments. I have no further major objections and I hope the work is well received. |