## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Risk of Harm in Synthetic and Biological Intervention Trials in Patients with Inflammatory Arthritis: Protocol for A Meta-Epidemiological Study Focusing on Contextual Factors |
| **AUTHORS** | Malm, Esben; Berg, Johannes; Nielsen, Sabrina; Ioannidis, John; Furst, Daniel; Smolen, Josef S; Taylor, Peter; Kristensen, Lars Erik; Tarp, Simon; Ellingsen, Torkell; Christensen, Robin |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Rachel Knevel<br>Leiden University Medical Center, Department of Rheumatology |
| **REVIEW RETURNED** | 25-Feb-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Clearly described research proposal. My main concern would be the lack of information published. Could this possible publication bias harm the analyses? |

| | |
|---|---|
| **REVIEWER** | Martin Stausholm<br>University of Bergen, Department of Global Public Health and Primary Care |
| **REVIEW RETURNED** | 08-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | The protocol is interesting and of high quality and therefore acceptable<br><br>Consider:<br>- Stating that two persons will select the trials, extract the data and judge the risk of bias in the strengths and limitations section<br>- Shortening the title so that is says ‘focusing on’ instead of ‘with focus on’<br>- In the section Study selection section replace ‘an RCT’ with ‘a RCT’<br><br>In table 1 it is stated that the comparison can be unclear. What do the authors mean by unclear comparison? |

| | |
|---|---|
| **REVIEWER** | Lewis Carpenter<br>King's College London, Health Psychology |
| **REVIEW RETURNED** | 21-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Many thanks for sending me this protocol for review. I have a few points of clarity before I can recommend for publication.<br><br>The protocol states that publications for inclusion of the review will be only taken from those included in other systematic reviews? Can the authors explain why only trials identified from systematic reviews will be sought? I would be much more in favour of a full systematic |

| | |
|---|---|
| | review of all literature regardless of whether it's been included in previous systematic reviews.<br><br>The authors stipulate that only trials since 2015 will be considered. Can the authors clarify if that means identified systematic reviews from which RCT will be identified, or does that include the RCT studies from the reviews themselves?<br>Can the authors also provide a rationale for this cut-point?<br><br>The power calculation is very vague and would benefit from more detail. It is not clear what analysis the power calculation is based on. Currently, the manuscript suggests 7,560 patients would provide sufficient power, but presumably this is at the patient level? The analysis alludes to looking at trial specific effects at the study level. How has power been determined for this main analysis?<br><br>Analysis mainly focused on univariable approaches (change in I2 and stratification). Is it important to consider factors in a multivariable manner to adjust for confounding effects? Can the authors comment on this.<br><br>In the aims and objectives section, I feel it would benefit from making it clearer what the specific outcomes of interest are. |

| | |
|---|---|
| **REVIEWER** | Dan McWilliams<br>University of Nottingham, ROD |
| **REVIEW RETURNED** | 03-May-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | bmj open 2021 049850<br>SLR protocol review<br><br>Clinical trials are generally very poor for estimating adverse events. In general, cohorts are superior for this. As most adverse events in RCTs are uncommon, I do not believe that they can be properly analysed by using population-level aggregate data. Individuals with adverse events may be atypical of the average study population. Surely an individual patient data meta-analysis is needed instead? My thoughts are that the level of precision will be very poor for estimating adverse events, and might also give misleading results (maybe in the opposite direction to the real effect).<br><br>Why are monoclonal antibodies and synthetic JAK-inhibitors included in the same review? They have quite different modes of action and the different biologic monoclonal antibodies also have different molecular targets. |

## VERSION 1 – AUTHOR RESPONSE

**Reviewer Reports:**

Reviewer: 1
Dr. Rachel Knevel, Leiden University Medical Center

Comments to the Author:
Clearly described research proposal.

Question 4: My main concern would be the lack of information published. Could this possible publication bias harm the analyses?

*Answer 4: We agree that this could be a risk. Trials published after 2004 are hopefully following the guidelines provided by the CONSORT Statement extension* on harms reporting. Further, we expect that there will be only minor reporting bias on our main outcomes; withdrawals and withdrawals due to adverse events and death, as this information is likely reported as part of the standard patient flow figure. However, we agree with the reviewer, as we expect that there (unfortunately) will be insufficient data on some of the contextual factors that we wish to investigate.*

* Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8

*Action 4: We have mentioned the possible limitation of missing data on contextual factors (Line 122-124).*

Reviewer: 2
Dr. Martin Stausholm, University of Bergen

Comments to the Author:
The protocol is interesting and of high quality and therefore acceptable.

Consider:

Question 5: Stating that two persons will select the trials, extract the data and judge the risk of bias in the strengths and limitations section

*Answer 5: We thank the reviewer for this suggestion, and we agree that it should be mentioned as a strength. Furthermore, we realize that the method section may be understood as each assessment will be done by both reviewers.*
*Action 5: We have mentioned the double study selection as a strength (line 120), and we have reformulated the description so it now should be clearer that each assessment will be done only once, by one of two reviewers (Line 324).*

Question 6: Shortening the title so that is says 'focusing on' instead of 'with focus on'

*Answer 6: Once again we thank the reviewer for the suggestion, and agree that it should be changed.*
*Action 6: We have changed the title (Line 4).*

Question 7: In the section Study selection section replace 'an RCT' with 'a RCT'

*Answer 7: We have investigated this further and recently asked a native English-speaking colleague, that that advised us to use 'an RCT', and explained that since the word RCT has a vowel sound, the preceding indefinite article should be 'an'. So, we hope that the reviewer will allow us to keep this way of spelling.*
*Action 7: No changes made (apologies).*

Question 8: In table 1 it is stated that the comparison can be unclear. What do the authors mean by unclear comparison?

*Answer 8: The comparator grouped as unclear is an ultimate last option (the least preferable). This is used when necessary, rather than excluding trials where the comparator arm is poorly described.*
*Action 8: Further elaborated in table 1.*

Reviewer: 3
Dr. Lewis Carpenter, King's College London

Comments to the Author:
Many thanks for sending me this protocol for review. I have a few points of clarity before I can recommend for publication.

Question 9: The protocol states that publications for inclusion of the review will be only taken from

those included in other systematic reviews? Can the authors explain why only trials identified from systematic reviews will be sought? I would be much more in favour of a full systematic review of all literature regardless of whether it's been included in previous systematic reviews.

*Answer 9: The reason for omitting a full systematic review is that we are likely to include the majority of trials, and since it is not at systematic review, but a meta-epidemiological study, a full systematic review is not required. Since we will include several large and thorough systematic reviews (including Cochrane reviews), we feel confident relying on their thorough literature searches and that the almost all eligible trials will be included. From a philosophy of science perspective the idea behind meta-research is that we perform (inductive) statistical inference from a representative sample of (almost) all trials; we will not make any claims on benefits and harms per se.*
*Action 9: No changes made.*

Question 10: The authors stipulate that only trials since 2015 will be considered. Can the authors clarify if that means identified systematic reviews from which RCT will be identified, or does that include the RCT studies from the reviews themselves?
Can the authors also provide a rationale for this cut-point?

*Answer 10: We thank the reviewer for pointing this out and realise that this part of the protocol should be more legible. Trials published before 2015 will indeed be considered, the limit is restricted to systematic reviews (non-Cochrane reviews). All trials included in any of the systematic reviews will be considered for eligibility.*
*Action 10: We have changed the description of the eligibility criteria (Line 216-218).*
*Additionally, we will make sure to clarify in text that the limit is restricted to systematic reviews not trials and visualise the process as a flowchart in the final reporting.*

Question 11: The power calculation is very vague and would benefit from more detail. It is not clear what analysis the power calculation is based on. Currently, the manuscript suggests 7,560 patients would provide sufficient power, but presumably this is at the patient level? The analysis alludes to looking at trial specific effects at the study level. How has power been determined for this main analysis?

*Answer 11: We agree completely with the reviewer and would like to add that it is generally well recognised that it is difficult to determine the anticipated power in meta-research per se. And Yes, this is based on patient level thus, we realize that this is inappropriate for our planned analyses. We thank the reviewer for picking this up.*
*Action 11: We have removed the sample size calculation and reformulated the section regarding sample size (line 385-393).*

Question 12: Analysis mainly focused on univariable approaches (change in I2 and stratification). Is it important to consider factors in a multivariable manner to adjust for confounding effects? Can the authors comment on this.

*Answer 12: We thank for this suggestion and it's something we have repeatedly discussed in the author group while drafting the protocol. We ended up choosing only univariable analyses among the prespecified analyses as the approach is exploratory and we do not have any hypothesis of which factors may be potential confounders of the results, and the number of analyses is already extensive. Further, we would worry that prespecifying multivariable analyses may limit the number of trials with sufficient data even more, as we already expect many of the contextual factors to be sparsely reported. We have added the point of meta-confounding to the limitations and introduced the possibility of performing (deconfounding) multivariable analyses.*
*Action 12: (line 121-124)+ (line 411-421).*

Question 13: In the aims and objectives section, I feel it would benefit from making it clearer what the specific outcomes of interest are.

*Answer 13: We thank the reviewer for pointing this out, and we agree that the aims and objective section could benefit from this change.*
*Action 13: The specific outcomes has been added to the aims and objectives section (Line 186-188).*

Reviewer: 4
Dr. Dan McWilliams, University of Nottingham

Comments to the Author:
Question 14: Clinical trials are generally very poor for estimating adverse events. In general, cohorts are superior for this. As most adverse events in RCTs are uncommon, I do not believe that they can be properly analysed by using population-level aggregate data. Individuals with adverse events may be atypical of the average study population. Surely an individual patient data meta-analysis is needed instead? My thoughts are that the level of precision will be very poor for estimating adverse events, and might also give misleading results (maybe in the opposite direction to the real effect).

*Answer 14: We thank the reviewer for this relevant concern, as it's something we have repeatedly discussed in the author group. In terms of our primary objective(s): We specifically wish to investigate associations between trial characteristics (trial level that is, not individual participant data level) harms in RCTs, and we hope that by including a large number of trials, we will be able to detect even some of the uncommon events. We are aware of the problem using population-level aggregated data and have now tried to elaborate a bit on this in the limitations (i.e. ecological fallacy). We would worry that we will not be able to get access to individual patient data due to GDPR regulations; other projects we are involved in clearly reveal some of the caveats (Holden et al; BMJ Open . 2017 Dec 22;7(12):e018971. doi: 10.1136/bmjopen-2017-018971). Instead, we hope the reviewer will support our vision, that by identifying important trial level contextual factors, harms may in the future be reported according to subgroups, and be summarised in future meta-analyses without needing independent patient data. These arguments are also important aspects that we are carrying forward in the OMERACT initiative, in the Contextual Factors Working Group (fyi: https://omeract.org/working-groups/contextual-factors/)*
*Action 14: Elaborated on the use of population-level aggregated data (line 124), and in the perspectives section (Line 444-446).*

Question 15: Why are monoclonal antibodies and synthetic JAK-inhibitors included in the same review? They have quite different modes of action and the different biologic monoclonal antibodies also have different molecular targets.

*Answer 15: Another good point and also something we have disused in the author group. As you suggest, studies of all the different treatments would be ideal, but the amount of scientific research (RCT's) is not large enough to make valid analyses this way. Rather we will explore whether the mode of action (e.g. Biologics vs small molecules) are different per se – and possibly adjust for that. We found that when combining all biological agents and small molecule treatments, the external validity of the inductive inference from the analysis would outweigh the disadvantage of combining different treatments. Action 15: no changes made.*

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Martin Stausholm<br>University of Bergen, Department of Global Public Health and Primary Care |
| **REVIEW RETURNED** | 11-Jun-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | It appears to be ready for publication |

| | |
|---|---|
| **REVIEWER** | Lewis Carpenter<br>King's College London, Health Psychology |
| **REVIEW RETURNED** | 07-Jun-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Many thanks for your considered responses.<br><br>Whilst the authors have provided a rationale for only identifying trials |

| | |
|---|---|
| | from other systematic reviews, I still have my concerns that potentially useful studies may be omitted, such as studies being excluded due to different inclusion/exclusion criteria or more recent trials not picked up when the search was conducted by the other reviews.<br>However I appreciate the authors argument that the analysis will be made on the basis of a representative sample of trials, so as long as this is made clear in the limitations if the authors are unwilling to reconsider a more formal systematic search.<br><br>Regarding the power analysis, I would be much more in favour of a sample size calculation. Whilst the authors are correct in stating that the number of trials will be fixed, they could still provide an estimate of what the effect is expected for the primary outcome (or indeed what effect would be considered clinically meaningful) and they have data on small sub-sample of studies to determine the rough sample size per study and an estimate of the heterogeneity between studies. This information would be sufficient to provide an estimate of the statistical power from the number of studies the authors anticipate. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer: 2

Mr. Martin Stausholm, University of Bergen

Comments to the Author:

It appears to be ready for publication

Reviewer: 3

Dr. Lewis Carpenter, King's College London

Comments to the Author:

Many thanks for your considered responses.

Question 6: Whilst the authors have provided a rationale for only identifying trials from other systematic reviews, I still have my concerns that potentially useful studies may be omitted, such as studies being excluded due to different inclusion/exclusion criteria or more recent trials not picked up when the search was conducted by the other reviews.

However I appreciate the authors argument that the analysis will be made on the basis of a representative sample of trials, so as long as this is made clear in the limitations if the authors are unwilling to reconsider a more formal systematic search.

Answer 6: We agree that this must be clear and have incorporated this in the discussion.

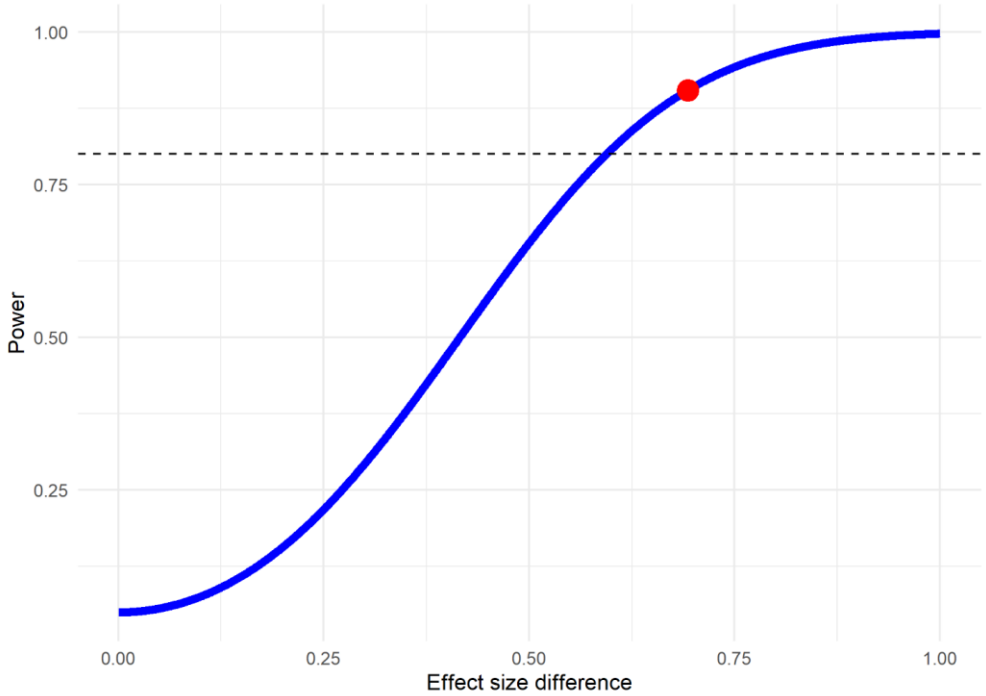
Action 6: We have added this as a limitation as part of the Discussion section (line 482-486)

Question 7: Regarding the power analysis, I would be much more in favour of a sample size calculation. Whilst the authors are correct in stating that the number of trials will be fixed, they could still provide an estimate of what the effect is expected for the primary outcome (or indeed what effect would be considered clinically meaningful) and they have data on small sub-sample of studies to determine the rough sample size per study and an estimate of the heterogeneity between studies. This information would be sufficient to provide an estimate of the statistical power from the number of studies the authors anticipate.

Answer 7: We regret that we did not succeed with the current omission. We obviously apologise for this inconvenience, and have now fully elaborated on this “power in meta-research” perspective.

Let us assume that we expect the first group to show an effect of log(OR) = log(1) = 0 with a standard error of 0.15, while the second group has an effect of log(OR) = log(2) = 0.69, and a standard error of 0.15. We use these assumptions as input in the power.analysis.subgroup()-function in R. In the output, we can see that the power of our imaginary subgroup test (90.4%) would be sufficient. The

output also tells us that, all else being equal, the effect size difference needs to be at least 0.595 in order to reach sufficient power.



Action 7: We have added a “Power and sample size considerations” section to the protocol (Line 424-443).

**VERSION 3 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Lewis Carpenter<br>King's College London, Health Psychology |
| **REVIEW RETURNED** | 02-Aug-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | My concerns have been addressed and I have no further comments. |