PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predictors of chronic pain and level of physical function in total knee arthroplasty: a protocol for a systematic review and meta-analysis
AUTHORS	Olsen, Unni; Lindberg, Maren; Denison, Eva Marie-Louise; Rose, Christopher; Gay, Caryl; Aamodt, Arild; Brox, Jens; Skare, Øystein; Furnes, Ove; Lee, Kathryn; Lerdal, Anners

VERSION 1 – REVIEW

REVIEWER	Jacek Kruczyński, Paweł Chodór Department of General Orthopedics, Orthopedic Oncology and Traumatology, University of Medical Sciences, Poznań, Poland
REVIEW RETURNED	14-Mar-2020

GENERAL COMMENTS	Study protocol for this meta-analysis encompasses everything that is required from a reliable and thorough research. High standards of Cochrane Handbook for Systematic Reviews and Interventions are fully met. That is crucial in a such complex and not entirely understood phentomenon as chronic pain after TKA. I look forward to seeing the results.

REVIEWER	Debbie Bean Auckland University of Technology New Zealand
REVIEW RETURNED	08-Apr-2020

GENERAL COMMENTS	Thank you for the opportunity to review this protocol for a systematic review and meta-analysis of predictors of chronic pain post-TKA. The topic is highly relevant and there are many studies which have been conducted which should be summarised. The methods described appear to me to be of the highest quality and potential problems in the review have been anticipated and a plan provided for how these will be dealt with. I look forward to reading the results of the systematic review. My only real questions were: There are other systematic reviews that have been conducted on the topic, some as recently as 2019, can the authors justify the need for a further review? Or describe the differences between previous reviews and this one more clearly? Also the plan is to include data from RCTs, and the authors plan to include data from 'the TKA arm'. I misunderstood at first, as there are RCTS which test interventions that are designed to prevent chronic pain in TKA. However as most of these studies would be aiming to alter risk factors I assume these are not the RCTs you're

	planning to include, can you describe more clearly the inclusion/exclusion of RCTs?

REVIEWER	Stephen Tregear Booz Allen Hamilton United States of America
REVIEW RETURNED	14-Apr-2020

GENERAL COMMENTS	Thank you for the opportunity to review this protocol. Methods and Analysis - Suggest the authors consider developing and including an analytic framework (see https://www.uspreventiveservicestaskforce.org/uspstf/uspstf-analytic-framework-child-and-adolescent-health-topics) to clearly lay out the potential key questions that could be addressed by the review, identify the outcomes that could be examined if data on the main outcomes is not available. In addition, the framework could be adapted to identify risk factors for increased pain and decreased functionality. Software - Authors propose to use STAT 16 for their meta-analysis - This is a good choice. Measures - The use of ORs and correlation coefficients as measures of association is acceptable. The authors propose to convert ORs and rs into a standardized metric, Hedges g - again an acceptable approach. Choice of metaanlytic model - Use of a random effects model in the light of expected heterogeneity is acceptable. The use of d the “overall correlation model” and the assumption that the within study correlations are unknown is acceptable. Assessment for publication bias - suitable methods for examining the potential for biased estimates of effect due to missing data are proposed. Assessment of heterogeneity - The authors propose to use I^2. This is an acceptable measure of heterogeneity and a threshold for "significant" heterogeneity is defined a priori. If heterogeneity is identified a series of exploratory analyses will be performed. These will include subgroup analyses and metaregression - both are acceptable methods but suggest that the authors define a priori the minimum number of studies that will be required per covariate included in metaregressions. This will show that that authors will be careful not to overfit the models. Proposed sensitivity analyses are sensible; however, the method for judging study quality (risk for bias) using QUIPS to classify a study into low risk, uncertain, and high risk for bias may be problematic. There is a vast literature on the relationships between quality as measured using a "scale" as apposed to the use of individual items on a checklist to examine the impact of potential for bias on outcome (for example - see Sander Greenland's articles from the 1980s. etc). Consider looking at impact of each item in QUIPS in the sensitivity analysis. Reporting - The proposed approach to reporting on results is appropriate with an effort being made to interpret the findings of the

	metaanalyses into more understandable terms. One item that seems to be missing though is mention of the role of the 95% CI as a measure of precision. Statements about the magnitude of association based on the point estimate alone without consideration of the precision of the estimate may be misleading. Summary - A well thought out protocol. I made some suggestions for consideration that I believe will make the protocol still stronger.

REVIEWER	Yohannes Woubishet Woldeamanuel Stanford University, USA Advanced Clinical & Research Center, Ethiopia
REVIEW RETURNED	24-Apr-2020

GENERAL COMMENTS	- This is an interesting protocol exploring predictors of chronic pain and level of physical function in total knee arthroplasty. This is important undertaking by the authors as TKA is increasingly becoming a common procedure with increasing life expectancy worldwide. I have the following comments. - Considering the different study designs and outcome measures that are planned to be combined, I recommend to include network meta-analysis as it will give a full picture of direct and indirect comparisons. - I suggest use of ROBv2 for risk of bias assessment. ROBv2 tool (Risk of Bias) is more commonly employed, and is more useful as it can assess risks related to missing data imputations and sensitivity analysis. - the protocol mentions OR (based on prevalence) as a potential measure of association. However, it has to be noted that OR is appropriate for retrospective, case-control, or some cross-sectional studies. For RCTs, relative risk or risk difference is the appropriate measuring, as it measures incidence compared to prevalence. OR for retrospective or cross-sectional studies. - Galbraith plot is more powerful display in meta-analysis, as it helps explore study precisions relationship with effect sizes, small-study bias as well as extent of heterogeneity. In addition, Baujat plot, Egger's regression plot, and exclusion sensitivity test need to be included as they demonstrate studies with largest heterogeneity. - Please include reference for heterogeneity cutoff you plan to use: "lower bound on the 95% CI on between-study I2 is greater than 50%". For example, the following article indicates 50% as moderate heterogeneity, and other cutoffs for different degrees of heterogeneity statistics. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. (2003) Measuring inconsistency in meta-analyses. British Medical Journal 327, 557-560 - missing data: please include assessment whether studies were compared between per-protocol and intent-to-treat analysis. And if so, what type of imputation was utilized for missing data handling. And whether there was sensitivity analysis to demonstrate whether results are robust post- to pre-imputation. - I suggest to include leave-one-out meta-analysis as part of sensitivity analysis, to demonstrate robustness of the results.

REVIEWER	YIMENG LIU University of Pittsburgh USA
REVIEW RETURNED	01-May-2020

GENERAL COMMENTS	In general, the author needs to provide more detail in how to extract the association measurement (I understand as the association measurement between each predictor and the two outcomes in the intervention arms with ATK). The current language description is confusing. I suggest adding some examples to help illustrate the data extraction process. Page 7 Line 35: There might be more than one intervention arms from the crossover study, please explain why you only extract intervention arm from the first period rather than the whole study. Page 7 Line 45: Estimand is not accurate used to describe "linear regression" and "correlation coefficient". These are the analysis methods. Page 7 Line 51: Please explain what type of the measurement will be extracted. For example, if there is a randomized control trial comparing the efficacy between the TKA arm and the placebo, the main analysis will compare the treatment and placebo. If there are predictors adjusted in the main model, do you extract the coefficient of the predictors in the model? However, the predictor effect will be same for treatment and placebo if there is no interaction term between the predictor and the treatment effect in the model. Suggest to give some examples to illustrate the measurement extracted. Page 8 Line 55: Will the meta-analysis be adjusted for the bias of each study? Page 9 Line 47: Hedges' g is a measure of effect size, if we extract the coefficient representing the predictor effect, is it calculated as the coefficient dividing its stand error? Page 9 Line 55: Is Hedges'g used as the outcome in the multivariate meta-analysis model? Page 10 Line 22, Line 44: Line 22 stated that the fixed effect analysis will not be compared with the random effect analysis which is not consistent with treating the fixed effect analysis as the sensitivity analysis (Line 44), please explain.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1. Study protocol for this meta-analysis encompasses everything that is required from a reliable and thorough research. High standards of Cochrane Handbook for Systematic Reviews and Interventions are fully met. That is crucial in such a complex and not entirely understood phenomenon as chronic pain after TKA. I look forward to seeing the results.
   Response: Thank you. We appreciate your comment.

Reviewer 2

1. There are other systematic reviews that have been conducted on the topic, some as recently as 2019, can the authors justify the need for a further review? Or describe the differences between previous reviews and this one more clearly?
   Response: We have revised the introduction to more thoroughly address the limitations of prior reviews and to more clearly justify the need for a new systematic review and meta-analysis (page 4, Introduction, section 3 and 4).

2. Also the plan is to include data from RCTs, and the authors plan to include data from 'the TKA arm'. I misunderstood at first, as there are RCTS which test interventions that are designed to prevent chronic pain in TKA. However as most of these studies would be aiming to alter risk factors I assume these are not the RCTs you're planning to include, can you describe more clearly the inclusion/exclusion of RCTs?
   Response: Thank you for pointing out that this was not clear. We have now added a more detailed description of the inclusion and exclusion criteria for RCTs, particularly regarding the intervention arm (page 5, Methods and analysis, Eligibility criteria).

Reviewer 3

1. Methods and analysis - Suggest the authors consider developing and including USPSTF analytic framework (see https://www.uspreventiveservicestaskforce.org/uspstf/uspstf-analytic-framework-child-and-adolescent-health-topics) to clearly lay out the potential key questions that could be addressed by the review, identify the outcomes that could be examined if data on the main outcomes is not available. In addition, the framework could be adapted to identify risk factors for increased pain and decreased functionality.
   Response: We appreciate this suggestion to include a framework. Although not mentioned in our originally submitted protocol, the Biopsychosocial model serves as the theoretical framework for this study. A statement about our use of the Biopsychosocial model has been added in the Methods section (page 5, Methods and analysis, section 2 and in Figure 1).

2. Measures - The use of ORs and correlation coefficients as measures of association is acceptable. The authors propose to convert ORs and rs into a standardized metric, Hedges g - again an acceptable approach.
   Response: Based on recent discussion on the Cochrane Statistical Methods Group mailing list (see the archives of that list for the second quarter of 2020) and feedback from reviewer 5, we have revised this aspect of our protocol. Briefly, we now plan to impute correlation coefficients (where necessary) and meta-analyze these via Fisher's z-transform. This approach is similar to a previous review by Lewis et al. (2015) on predictors of post-TKA pain as mentioned in the article (page 8, Methods and analysis, Measures of association).

3. Regarding subgroup analyses and meta-regression, and overfitting of models - define a priori the minimum number of studies that will be required per covariate included in meta-regressions.
   Response: Thank you for this suggestion. We have modified the protocol to pre-specify the minimum number of studies required per covariate for exploratory meta-regression (page 10, Methods and analysis, Subgroup analysis).

4. Proposed sensitivity analyses are sensible; however, the method for judging study quality (risk for bias) using QUIPS to classify a study into low risk, uncertain, and high risk for bias may be problematic. There is a vast literature on the relationships between quality as measured using a "scale" as opposed to the use of individual items on a checklist to examine the impact of potential for bias on outcome (for example - see Sander Greenland's articles from the 1980s. etc). Consider looking at impact of each item in QUIPS in the sensitivity analysis.
   Response: As suggested, we have revised the protocol to perform sensitivity analyses with respect to each domain of the QUIPS tool regression (page 10, Methods and analysis, Sensitivity analysis).

5. Reporting - The proposed approach to reporting on results is appropriate with an effort being made to interpret the findings of the meta-analyses into more understandable terms. One item that seems to be missing though is mention of the role of the 95% CI as a measure of precision. Statements about the magnitude of association based on the point estimate alone without consideration of the precision of the estimate may be misleading.
   Response: We agree that it would be misleading to present point estimates without also quantifying their precision. The "Presentation and interpretation of results" section (page 10) states that we will present 95% CIs, and we will do so for all estimates.

Reviewer 4

1. Considering the different study designs and outcome measures that are planned to be combined, I recommend to include network meta-analysis to give a full picture of direct and indirect comparisons.
   Response: We appreciate this reviewer's review, but based on this and their other comments, we suspect they think we are planning a meta-analysis of the effects of an intervention, rather than our plan to review the associations between factors such as age and BMI and post-surgical pain (i.e., we do not plan to compare any interventions). Network meta-analysis (NMA) is used to compare multiple interventions, and is therefore not the correct tool for our research question. However, it is possible to implement NMA using the multivariate meta-analysis method we have chosen, so the two methods are very closely related.

2. I suggest use of ROBv2 for risk of bias assessment. ROBv2 tool (Risk of Bias) is more commonly employed, and is more useful as it can assess risks related to missing data imputations and sensitivity analysis.
   Response: ROBv2 is an excellent tool for assessing risk of bias in randomized trials of interventions. However, as stated above, we plan to study factors associated with pain after surgery, so ROBv2 is not appropriate (because there are no interventions to compare). We did a scoping exercise when writing the protocol, which revealed that most of the studies we are likely to include are non-randomized. If we were studying interventions, we would therefore consider the ROBINS-I tool, which is a variant of the ROBv2 tool, but for non-randomized studies of interventions. Because we are studying associations, we identified and plan to use the QUIPS tool, which was developed for assessing risk of bias in prognostic factors research (Hayden et al. 2013. Assessing Bias in Studies of Prognostic factors.). QUIPS is also the tool recommended by the Cochrane Methods Prognosis group for this type of question.

3. The protocol mentions OR (based on prevalence) as a potential measure of association. However, OR is appropriate for retrospective, case-control, or some cross-sectional studies. For RCTs, relative risk or risk difference is the appropriate measuring, as it measures incidence compared to prevalence. OR for retrospective or cross-sectional studies.
   Response: Because we are reliant on estimates of association published in included studies, we cannot choose how the authors of those studies analyzed their data or presented results. We are therefore likely to have ORs, RRs, correlation coefficients, and linear regression coefficients (as well as SEs or CIs for those). Briefly, our plan is to adhere to standard meta-analysis practice and perform analysis on a common scale, imputing from other scales as necessary (see our response on this issue to reviewer 3). Methods exist for imputing between ORs, RRs, regression coefficients, and correlation coefficients (this literature goes back to Pearson 1900). We note that our inclusion criteria do not include retrospective, case-control, or cross-sectional studies.

4. Galbraith plot is more powerful display in meta-analysis, as it helps explore study precisions relationship with effect sizes, small-study bias as well as extent of heterogeneity. In addition,

Baujat plot, Egger's regression plot, and exclusion sensitivity test need to be included as they demonstrate studies with largest heterogeneity.
Response: Thank you for these suggestions. We are aware that there are multiple ways of presenting and analyzing relationships between magnitude of effect and precision. We prefer to follow Cochrane Handbook recommendations and practice, which is to present funnel plots (see Page, Higgins, Sterne, Chapter 13 of version 6 of the Cochrane Handbook), because readers are likely to be more familiar with these plots. However, we are aware that funnel plots (and indeed the alternative plots the reviewer mentions) have limitations.

5. Please include reference for heterogeneity cutoff.
   Response: We have now clarified how we will interpret I^2 and have provided a reference as requested (page 10, methods and analysis, Subgroup analysis).

6. Missing data: Please include assessment whether studies were compared between per-protocol and intent-to-treat analysis. And if so, what type of imputation was utilized for missing data handling. And whether there was sensitivity analysis to demonstrate whether results are robust post- to pre-imputation.
   Response: Regarding intention to treat and per-protocol analysis, as noted above, the reviewer may be thinking in terms of comparing treatments, not about associations between factors such as age, BMI, and outcomes such as pain. The intention-to-treat principle (ITT) is applicable when there are one or more treatment comparisons and some participants are assigned to receive the treatment and others are assigned to one or more comparators (as in the case of RCTs). An ITT analysis studies the effect of the “policy” to treat, while per-protocol studies the effect of actually treating. Our research question does not concern comparisons of treatments, so these concepts as we understand them do not apply.
   However, we do not plan to perform our own imputation because we think it is currently impractical to try to pre-specify and then use imputation methods that can be applied in a multivariate setting in which studies can report a wide range of measures of association (however, if the reviewers know of such methods that are practical to use in this setting, we would appreciate their suggestions). Instead, we plan to address the issue of missing data via use of the QUIPS tool (see Hayden et al. 2013. Assessing Bias in Studies of Prognostic factors) and the sensitivity analyses we have now added based on that tool (see our response to reviewer 3, comment 4). Specifically, see all items under domain 2 and items 1a, 1d, 1f, and 5e of QUIPS.

7. I suggest to include leave-one-out meta-analysis as part of sensitivity analysis, to demonstrate robustness of the results.
   Response: We thank the reviewer for this suggestion and have added a leave-one-out sensitivity analysis to the protocol (page 10, Methods and analysis, Sensitivity analysis).

Reviewer 5.

1. In general, the author needs to provide more detail in how to extract the association measurement (I understand as the association measurement between each predictor and the two outcomes in the intervention arms with ATK). The current language description is confusing. I suggest adding some examples to help illustrate the data extraction process.
   Response: We have rewritten some of the text that relates to this issue and hope that it is now clearer (page 7-8, Methods and analysis, Study selection and data extraction).
2. Page 7 Line 35: There might be more than one intervention arms from the crossover study, please explain why you only extract intervention arm from the first period rather than the whole study.
   Response: An explanation for this decision (to avoid carry-over effects) has been added (page 5, Methods and analysis, Eligibility criteria).

3. Page 7 Line 45: Estimand is not accurate used to describe "linear regression" and "correlation coefficient". These are the analysis methods.
   Response: We thank the reviewer for pointing out this error and have changed “estimand” to “analysis type” (page 5, Methods and analysis, last row of Table 3).

4. Page 7 Line 51: Please explain what type of the measurement will be extracted. For example, if there is a randomized control trial comparing the efficacy between the TKA arm and the placebo, the main analysis will compare the treatment and placebo. If there are predictors adjusted in the main model, do you extract the coefficient of the predictors in the model? However, the predictor effect will be same for treatment and placebo if there is no interaction term between the predictor and the treatment effect in the model. Suggest to give some examples to illustrate the measurement extracted.
   Response: We apologize that this was not clear. Only associations within the TKA arm of an RCT would be extracted (page 5, Methods and Analysis, Eligibility criteria). In the example provided by the reviewer, if the RCT only included an analysis that included both TKA and placebo arms, and did not include association’s specific to the TKA arm, no measurements would be extracted. Even if the final model included no interaction between the predictor and the treatment effect, we would not be able to conclude that the predictor coefficients would be the same for both arms, and therefore we would not be able to extract them for use in this meta-analysis. However, we would contact study authors to try and obtain arm-wise results for inclusion in the meta-analysis as stated (page 8, Study selection and data extraction).
5. Page 8 Line 55: Will the meta-analysis be adjusted for the bias of each study?
   Response: We do not plan to perform any statistical adjustment for risk of bias, but we will interpret our results in light of our judgements about risk, and we have now added a sensitivity analysis that looks at each domain of the QUIPS tool (page 10, Methods and analysis, Sensitivity analysis). One of the authors is currently working on a bias-adjusted network meta-analysis. This kind of analysis is nontrivial because it is generally necessary to judge the likely direction of the bias (i.e., a result at high risk of bias may be biased upwards, while another result at the same risk of bias may be biased downwards). The literature and practice in this regard seem to be relatively immature. For example, the new ROBv2 tool from Cochrane (which is not yet implemented across Cochrane) asks raters to optionally judge likely direction of bias. Our understanding of this is that extraction of such data is anticipated to be useful in the future when methods for bias adjustment are more mature. We suspect that bias-adjusted meta-analysis is a fruitful area for methodological research, but given that it is not routinely done in Cochrane reviews that use simple pairwise meta-analyses, we are reluctant to apply it in a much more complex setting.

6. Page 9 Line 47: Hedges’ g is a measure of effect size, if we extract the coefficient representing the predictor effect, is it calculated as the coefficient dividing its stand error?
   Response: We very much appreciate this comment, which contributed in part to us rethinking the scale on which to perform meta-regression. We have rewritten this section of the protocol (page 9, Methods and analysis, Assessment of non-reporting bias).

7. Page 9 Line 55: Is Hedges’g used as the outcome in the multivariate meta-analysis model?
   Response: Please see our response to the previous comment.

8. Page 10 Line 22, Line 44: Line 22 stated that the fixed effect analysis will not be compared with the random effect analysis which is not consistent with treating the fixed effect analysis as the sensitivity analysis (Line 44), please explain.
   Response: We agree that this was confusing and have modified the protocol to remove the fixed effects analysis under the section “sensitivity analysis”. Because we will potentially have substantial heterogeneity, a fixed effects model is probably inappropriate anyway. The random effects model essentially collapses to the fixed effects model if there is no heterogeneity, rendering a separate fixed effects analysis somewhat meaningless.

VERSION 2 – REVIEW

REVIEWER	Yohannes Woubishet Woldeamanuel Stanford University, USA Advanced Clinical & Research Center, Ethiopia
REVIEW RETURNED	02-Jul-2020

GENERAL COMMENTS	My comments are addressed.